DOCUMENT RESUME

ED 273 197 HE 019 611

AUTHOR	Adelman, Clifford, Ed.
TITLE	Assessment in American Higher Education. Issues and Contexts.
INSTITUTION	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	OR-86-301
PUB DATE	86
NOTE	90p.; Commissioned documents prepared for the National Conference on Assessment in Higher Education (Columbia, SC, October 1985) sponsored by the U.S. Office of Educational Research and Improvement.
AVAILABLE FROM	U.S. Department of Education, Office of Educational Research and Improvement, Special Services Division, Publications Branch, 80 F Street, N.W., Washington, DC 20208.
PUB TYPE	Collected Works - Conference Proceedings (021) -- Viewpoints (120)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Academic Achievement; Accountability; Business Administration Education; *College Students; Cost Estimates; *Educational Assessment; Educational Policy; Educational Quality; *Educational Testing; Engineering Education; *Higher Education; Institutional Characteristics; *Outcomes of Education; Preservice Teacher Education; Professional Education; Program Evaluation; *Student Evaluation; Test Selection

ABSTRACT

Emerging and traditional forms of assessment in U.S. higher education are considered in this collection of five conference papers from the 1985 National Conference on Assessment in Higher Education. Included are a foreword by Education Secretary William J. Bennett and concluding remarks by Clifford Adelman. Paper titles and authors are as follows: "The Growing Interest in Measuring the Educational Achievement of College Students" (Terry W. Hartle); "Assessing Outcomes in Higher Education" (John Harris); "The Costs of Assessment" (Peter T. Ewell, Dennis P. Jones); "Assessment in Higher Education: To Serve the Learner" (Georgine Loacker, Lucy Cromwell, Kathleen O'Brien); and "Assessment in Career-Oriented Education" (Sandra E. Elman, Ernest A. Lynton). The papers include descriptions of different testing programs and the sponsoring organization's address and phone number. Assessment programs and cost estimates for four types of colleges are also covered: a small, private liberal arts college; a major public research university; a regional, comprehensive, public university; and a mid-sized community college. The conclusion briefly considers the role of judgment in culture and language as a theoretical ground for thinking about assessment. Key issues implicit in the papers are also identified, and a 48-item bibliography of technical and theoretical references is provided. (SW)

ED273197

Assessment in American Higher Education

Issues and Contexts

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

HE019 611

Essays by
Terry W. Hartle
John Harris
Peter T. Ewell and Dennis P. Jones
Georgine Loacker et al.
Sandra Elman and Ernest Lynton

Edited with a Conclusion by
Clifford Adelman

Foreword by
William J. Bennett
Secretary of Education

OERI

HE

ASSESSMENT IN AMERICAN HIGHER EDUCATION

Issues and Contexts

Essays by
Terry W. Hartle
John Harris
Peter T. Ewell and Dennis P. Jones
Georgine Loacker et al.
Sandra Elman and Ernest Lynton

Foreword by
William J. Bennett
Secretary of Education

Edited with a Conclusion by
Clifford Adelman

Office of Educational Research and Improvement
U.S. Department of Education

Production Services by
The American Association for Higher Education

For sale by the Superintendent of Documents, U.S. Government Printing Office Washington, D.C. 20402

CONTENTS

Foreword	William J. Bennett	i
About this Volume		v
About the Authors		vii
The Growing Interest in Measuring the Educational Achievement of College Students	Terry W. Hartle	1
Assessing Outcomes in Higher Education	John Harris	13
The Costs of Assessment	Peter T. Ewell and Dennis P. Jones	33
Assessment in Higher Education: To Serve the Learner	Georgine Loacker, Lucy Cromwell, and Kathleen O'Brien	47
Assessment in Career-Oriented Education	Sandra E. Elman and Ernest A. Lynton	63
To Imagine an Adverb: Concluding Notes to Adversaries and Enthusiasts	Clifford Adelman	73

FOREWORD

Our nation has created the world's finest system of higher education. At its best, it combines the best research and teaching with the greatest variety of educational programs available anywhere. It is a system composed of universities, colleges, junior colleges, and professional schools of almost every description. Together they provide our citizens with multiple opportunities to tailor an educational program to their changing goals and circumstances throughout life.

Today, 50 percent of American high school graduates go on to enroll in postsecondary institutions, with total enrollments at almost 18 million. Expenditures by these institutions have nearly doubled since 1966; they totalled $90 billion in 1984. Funding from federal, state, and local governments accounted for almost half this total—$44 billion in 1984, up to $26 billion in 1966 when adjusted for inflation. The private sector has also provided substantial—and steadily increasing—support for higher education.

The American people have been generous to our colleges and universities and this generosity derives from the belief that these institutions are an indispensable foundation of our economic progress and national well-being. It rests on the firm belief that these institutions offer a gateway to the American dream. Given this importance we ascribe to higher education, as well as its growing costs, it is only reasonable that students, parents, government officials, and others should look for—and expect to find—evidence that they are getting their money's worth. This is a particularly important matter for students from less financially fortunate homes, students for whom higher education may be a crucial avenue to success.

Many students now receive an excellent education from our institutions of higher education. But the health and vitality of these institutions depend upon the creation and maintenance of rigorous standards of achievement for students, faculty members, and institutions themselves. There is wide agreement that the quality of undergraduate liberal arts education at a number of colleges and universities is not what it should be. We have all heard reports that many of our graduates do not possess the knowledge, skills, or, in some cases, the civic virtues of a well-educated person. Some evidence is fragmentary, anecdotal, or impressionistic; other indicators are more tangible: student performance declined in 11 of 15 major Subject Area Tests of the Graduate Record Examination between 1964 and 1982.

We have seen five major reports in just over one year that have been critical of various aspects of undergraduate education. These reports contain some troubling findings. For example, a 1984-85 survey by the American Council on Education indicates that a student can obtain a bachelor's degree from 72 percent of all American colleges and universities without having studied American literature and history; from 75 percent without having studied European history; and from 86 percent without having studied the civilizations of classical Greece and Rome. The Modern Language Association reports that, in 1966, 89 percent of all institutions required foreign language study for the bachelor's degree; this dropped to 53 percent in 1975, and to 47 percent in 1983.

As the recent Association of American Colleges report, *Integrity in the College Curriculum*, states, higher education has gone through a period in which there seemed to be more confidence "about the length of college education than its content and purposes." The neglect of the real purposes and goals of education strikes at the very integrity of higher education.

I am encouraged by signs that our colleges and universities are now recognizing the need to improve the quality of undergraduate education. For, while construed by some as an indictment of higher education, these reports are, in fact, a promising sign. They have recognized the danger of declining quality and provided guidance on how the problems can be overcome. These reports are, for the most part, products of the academy. They are *by* its members *to* its members, and it is the members of the academy who must take the lead to solve these problems.

The quality of the "product"—of the education actually received—is the central issue. From the perspective of society at large, the worrisome inadequacies are inadequacies not so much of processes as of *outcome* and *performance*. At the undergraduate level, we might—at the risk of oversimplifying—state the fundamental problem thus: We are uncertain what we think our students should learn, how best to teach it to them, and how to be sure when they have learned it.

Stated this way, the criticisms don't sound so different from the dominant criticisms of secondary education these past few years. Of course, the college and the high school have differences as well as similarities. Nevertheless, I believe that higher education could learn a lesson from the reform movement taking place at the elementary and secondary level.

For one, the call for assessment has been good for elementary and secondary education. In what is now called "effective schools research," scholars have been successful in examining schools that appear to produce good students and then identifying those institutions' common characteristics. And, as it turns out, among the characteristics of effective schools is a willingness to define educational goals, to assess performance in meeting those goals, and to make the results of such assessments available to the community. Institutions of higher education should do the same.

I believe that the quality of higher education must be improved, but I also believe that the primary force for that improvement should come from the institutions themselves. Our colleges and universities must do a better job of providing a coherent and rigorous curriculum for students. They must do a more conscientious job of stating their goals, of gauging their own success in relation to those goals, and of making their results available to everyone—students, prospective students, parents, citizens, and taxpayers.

Apart from the essential skills and fundamental knowledge that we expect all colleges and universities to impart, there are individual institutional goals that vary enormously from campus to campus. It is only sensible that each school appraise its own progress toward its particular goals. This is the surest way to turn the lofty statements of college catalogues into actual classroom practice. If we are to keep our promises to students, we must be willing to honestly assess our strengths—and our shortcomings. Such acknowledgement is the surest way to maintain institutional integrity; it is also the best way to maintain institutional sovereignty and self-government.

This volume is intended to assist those striving to develop and carry out better means of assessment. The papers collected here summarize recent trends in assessment and describe a number of promising institutional efforts. This research reveals that some institutions of higher education are beginning to assess student outcomes more rigorously as a means of assessing learning. While their methods vary, some colleges and universities are beginning to set competency levels in certain content areas that must be met before a student can be promoted.

This research also shows that the concept of assessment extends to many different methods—standardized tests, interviews, questionnaires, reviews of students' written work over four years, reviews of extra-curricular activity, studies of alumni and dropouts, surveys of students' use of time, surveys of graduates' use of time and more. Some results can be expressed in numerical terms; many obviously cannot. But no matter what the form, judgments need to be made so that institutions can assure the public and themselves that they are doing what they say they are doing.

Some argue that no matter what form assessment assumes, it is bound to damage teaching. Some fear that assessment is certain to lead to the practice called "teaching to the test." This, I believe, is an argument that tries to put the cart before the horse. What does an institution want to assess? It wants them to learn the ideas, the thoughts, the works, the skills and methods that the faculty, department, college, and university believe an educated person should possess. The institution must set its own goals, it must articulate a vision, it must delineate standards, and then it is quite all right to teach to those goals and standards. When a college or university does that, it does nothing shameful. It simply does what it set out to do, and then checks to see how well it has succeeded.

If assessment is done right, if it is done with care, it is nothing more than a means to measure whether students are learning what the college says they should learn (and that which it usually boasts they will learn). Any test, therefore, must be designed to fit standards and goals for which the institution aims. And it may not even look like a "test." Set standards first, articulate the vision of the educated person first, then formulate the method of assessment. If it's done in the right order, there's no reason to fear "teaching to the test." What you will get is teaching to a vision of an educated human being. And that's exactly what we should want.

Some skeptics might say: But those goals of which you speak, the qualities that make an educated man or woman, are qualities no one can accurately measure. As William James said, the best that a college education can aspire to accomplish is to help you know a good man when you see him. It is the intangibles that lie at the heart of higher education. And if you try to deny this, the skeptics might tell me, then we will bring to witness your own words.

Remember, the skeptics might add, what you wrote at the National Endowment for the Humanities in your own report on higher education. You wrote that students would "grapple with life's enduring, fundamental questions: What is justice? What should be loved? What deserves to be defended? What is courage? What is noble? What is base? Why do civilizations flourish? Why do they decline? . . . What can I know? What may I hope for? What is man?"

Indeed, these are some of the things that matter the most in higher education. Can we assess learning when it comes to these things? Yes, I believe we can, if students are given the chance to say what they know and how they've been affected by that knowledge. There is no reason why we can't ask students broad questions and assess the depth of their answers. As a teacher I did it all the time.

In fact, I believe that thoughtful assessment will bear out the truth of what I have been saying about the matters that lie at the heart of higher education. I believe we will find that students regard their college experience as more valuable if they have been required to confront the truly great issues, great thoughts, and great writers. Real assessment, I think, will bring support for these themes for which I have argued in the past. It will give students a chance to tell us what has mattered to them. Thus we can judge their enterprise as well as our own.

I am optimistic that our colleges and universities will turn the zeal for reform to their own advantages—to all of our advantage. We at the Department of Education are trying to help. The federal government cannot and should not play the primary role in the assessment of higher education. But we are interested in getting behind good ideas where we can. We are interested in fostering good ideas and I believe this volume contains a number of them. I hope it will stimulate still more and that the ensuing creation of more effective structures of assessment will help us meet the important challenges facing higher education.

William J. Bennett
Secretary of Education

About this Volume

The five papers in this collection were selected from a variety of commissioned documents prepared for a National Conference on Assessment in Higher Education sponsored by the Office of Educational Research and Improvement of the U.S. Department of Education, designed by the American Association for Higher Education, and hosted by the University of South Carolina in Columbia in October 1985.

The Conference was one of a series of dissemination activities conducted during the year following the Department's release of *Involvement in Learning: Realizing the Potential of American Higher Education*, the national report that raised assessment to a first principle of improvement in higher education. The intention of this particular conference was to provide a series of introductions to the current impetus, politics, uses, and general methodologies of assessment. In our customary language of curriculum, these papers are thus selections from the "General Education" portion of the field; and the collection as a whole is not meant to be comprehensive.

Indeed, as Secretary Bennett's "Foreword" implies, many institutions of higher education are just starting out on the long road of developing "effective structures of assessment." It is partly for this reason that the editor offers a concluding essay indicating the technical questions and issues that must be addressed as those institutions move from introductory to advanced study.

Acknowledgements

Some special thanks are in order. Not only for this volume (which is the fifth and last of a series following up on the major themes and background work of *Involvement in Learning*), but for all our efforts to promote discussion of the recent recommendations for reform in American higher education, we owe much to the guidance, wisdom, and just plain hard work of Theodore Marchese, vice president of the American Association for Higher Education. The publication series itself, a major force in those discussions, would not have been possible without the careful editorial work and design selections of Jann Teeple-Hewes at AAHE. Within the Department of Education, support for these publications was in no small measure due to the enthusiasm and advocacy of Sharon Horn and John Wirt of the Office of Educational Research and Improvement, and, from the moment he arrived in the fall of 1985, Assistant Secretary Chester E. Finn, Jr. As Secretary Bennett notes in his Foreword, the Department stands ready not only "to get behind good ideas" but also "in fostering good ideas," and the support of my colleagues in this undertaking certainly bears him out.

About the Authors

TERRY W. HARTLE is a resident fellow at the American Enterprise Institute for Public Policy Research (AEI) in Washington, DC. He is coauthor of two recent volumes published by AEI: *Excellence in Education: The States Take Charge* and *The Great American Job Machine*.

JOHN HARRIS is E. M. Ijams professor of Christian Education at David Lipscomb College in Nashville, Tennessee. He is also a member of the National Advisory Committee on Accreditation and Institutional Eligibility of the U.S. Department of Education.

PETER T. EWELL is senior associate at the National Center for Higher Education Management Systems in Boulder, Colorado. He has written five books and monographs on assessment; his most recent book, *The Self-Regarding Institution* (1984), has been critically acclaimed for its contributions in this area.

DENNIS P. JONES is president of the National Center for Higher Education Management Systems. He is the author of two books and has written extensively on the development of data bases in colleges and universities, information for management decision making, facilities planning, and state-level budget formulas.

GEORGINE LOACKER is professor of English at Alverno College in Milwaukee, Wisconsin. She also chairs Alverno's Assessment Council and the Analysis/Communication Division and has written extensively on assessment at Alverno.

LUCY CROMWELL is associate professor of English at Alverno College in Milwaukee. She coordinates Alverno's Analysis Department and directs a FIPSE-funded project on critical thinking.

KATHLEEN O'BRIEN is assistant professor of management at Alverno College and chairs the Problem Solving/Social Interaction Division. She also heads a project on leadership development sponsored by the National Association for Bank Women and the National Executive Service Corp.

SANDRA E. ELMAN is senior associate at the McCormack Institute of Public Affairs at the University of Massachusetts, Boston. She is coauthor of *Professional Service and Faculty Rewards: Toward an Integrated Structure*

ERNEST A. LYNTON is commonwealth professor and senior associate at the McCormack Institute of Public Affairs at the University of Massachusetts, Boston. He and Sandra Elman have authored a book on the role and functions of the university, tentatively titled *The Extended University*.

CLIFFORD ADELMAN is senior associate in the Office of Educational Research and Improvement, U.S. Department of Education. Before joining the Department in 1980, he taught at City College of New York and Yale University and was associate dean and assistant vice president for academic affairs at William Paterson College of New Jersey.

NOTE: The views expressed in the five essays and in the conclusion are those of the authors and do not represent the opinions or positions of the United States Department of Education.

The Growing Interest in Measuring the Educational Achievement of College Students

by Terry W. Hartle

The ground is shifting on American colleges and universities. After two decades of focusing on issues of equal opportunity and student access, the emphasis is increasingly on educational quality and the intellectual skills of students. One recent report on higher education bluntly warned: "the quality and meaning of undergraduate education has fallen to a point at which mere access has lost much of its value" (Southern Regional Education Board, 1985).

There is no shortage of evidence that academic quality needs some attention:

- A large number of college students need remediation. Research suggests that the average community college freshman is reading at an eighth grade level.[1]
- Student performance on the verbal section of tests of general learned abilities (such as the Graduate Record Examination) has declined sharply in the last decade. Performance on some professional licensing exams, such as state bar examinations, has also fallen.[2]
- State policy makers have begun to raise questions about the nature and quality of instruction at public colleges and universities.[3]
- Faculty members overwhelmingly believe that today's students have less interest in learning than those they taught at the outset of their careers.[4]
- Sharp criticisms of higher education have begun to appear in popular magazines, accusing colleges of everything from poor students to no quality control to price gouging.[5]

In the last two years, major reports from diverse groups have described these problems in detail and issued strong calls for improvements in academic programs. In *Involvement in Learning* (1984), the Study Group on the Conditions of Excellence in American Higher Education recommended a systematic program to assess the knowledge, capacities, and skills developed in students by academic and co-curricular programs. William Bennett, then chairman of the National Endowment for the Humanities, issued a statement, *To Reclaim A Legacy* (1984), that called for renewed attention to the humanities and urged college and university presidents to take a leading role in curricular reform. The Association of American College's report, *Integrity in the College Curriculum* (1985), referred to the absence of institutional accountability as "one of the most remarkable and scandalous aspects" of higher education and proposed that college faculties design and monitor appropriate techniques for measuring student progress.

Most recently, the Southern Regional Education Board's Commission for Educational Quality (1985) called for the establishment of a "new covenant" involving the public, its political representatives, and higher education, to find ways to improving academic quality while maintaining student access. Such a goal, the Commission concluded, will require new measures of student performance.

There has already been some movement to address quality concerns. Many colleges have revised their curricula and others are considering changes. A number of institutions have tightened their admissions requirements hoping to insure that students enter with a greater level of knowledge and preparation. Some institutions have begun to use commercially developed products to measure student progress and achievement while in college.

More promising (or ominous, depending upon your perspective), are the efforts of some state governments to increase educational quality at public institutions. A recent study by the College Board (1985) found that twenty-four states now set minimum admissions requirements for freshmen at all public institutions within their borders. Sixteen of these states have enacted, or are considering, more stringent admissions policies.

Other state actions include mandating achievement tests and revising funding formulas to reward colleges that demonstrate gains in student learning.

If the calls for change and the actions taken so far have a common theme, it is a desire to assure higher levels of student performance. Much of the public discussion seems focused on the outcomes of a postsecondary education, and proposals for better assessment of student learning are common. Assessment is a neutral enough word and it carries little of the negative baggage that other phrases (e.g., accountability testing) would bring along. But assessment has a number of different meanings, and is rapidly becoming an overused word that means different things to different people in different settings.

This paper seeks to provide an overview of the current interest in assessment: what it is and what it means in higher education, how it is being pursued, the questions it raises, and its future. The intention is not to answer questions as much as to raise them, in hopes that the other papers in this volume will shed more light on the host of issues that merit attention.

What Is Assessment and What Does It Mean in Higher Education?

The theory of assessment began to emerge in the late 1930s, thanks to the research of Henry A. Murray and his colleagues at the Harvard Psychological Clinic. The first large-scale effort to put assessment into practice was made by the Office of Strategic Services (OSS) during the Second World War to evaluate candidates for especially dangerous jobs. In the mid 1960s, Douglas Bray extended the assessment method into corporate settings by starting a long-term study of a group of new managers at AT&T and following their development. A decade later, assessment centers were relatively common in the corporate world; MacKinnon (1975) estimated that there were as many as 1,000 of them.

In education, assessment is often used interchangeably with testing, evaluation, and/or measurement. It is different from them in important respects, but drawing the distinctions is often difficult. Assessment is derived from a Latin word meaning "to sit beside" or "assist in the office of the judge." Thus, the word refers to the gathering and assembling of data into an interpretable form. The evidence is focused on the individual subject, or "assessee." MacKinnon's definition (1975) of the traditional meaning of assessment is a good one:

> . . . assessment is a method for the psychological evaluation of individuals that involves testing and observing individuals in a group setting, with a multiplicity of tests and procedures, by a number of staff members. Through a pooling of test scores and subjective impressions, the assessors formulate psychodynamic descriptions of the assessed subjects which, hopefully, will permit prediction of the assessees' behavior in certain kinds of roles and situations.

The *Encyclopedia of Educational Evaluation* emphasizes that assessment is a "multitrait-multimethod" technique, meaning that it involves a number of variables (rather than a single measurement such as a test) and uses a number of different procedures to measure them. Its techniques may also involve multiple sources (data on the same variable is collected from different sources) and/or multiple judges (a number of assessors may interpret the evidence and make judgments).[6]

Meeting all these criteria is difficult. The best known educational "assessment," the National Assessment of Educational Progress (NAEP), for example, meets some of these requirements, but not all. It tests school children in different age groups in several academic areas using different techniques (e.g., multiple choice, essay). The evidence allows analysts to make judgments about education quality for large segments of the population. But individual scores are not issued; the data are aggregated before analysis, interpretation, and reporting. A true assessment would focus on the individual learner.

Within higher education, the situation is even more complicated; "assessment" sometimes refers to half a dozen separate (but related) activities. The first, which comes closest to the historic meaning, involves multiple measures and observers to track intellectual and personal growth over an extended period of time. The best, and perhaps only, comprehensive example of this approach is Alverno College. Over the course of a four-year career, the typical student will undergo more than 100 performance assessments. The college uses simulations that require students to demonstrate one or more of eight core abilities and six levels of performance within them. Criteria for evaluating the abilities remain the same for all disciplines. It is a multiple-judge approach: faculty, peers, community members, professionals, and others may act as assessors. Research done at Alverno suggests that the approach has been very effective.[7]

The second, and perhaps most common, meaning of assessment in higher education refers to state-mandated requirements to evaluate students and/or academic programs. Some states use the pass rates on professional licensing examinations as indicators of quality, especially in teaching and nursing. Other states use testing for counseling and placement. Still others employ tests as a promotional gate that students must pass through before receiving their degree or moving on toward further education.

Postsecondary student testing has risen in popularity partly because of the states' experience with minimum competency tests for elementary and secondary school students. A decade ago, few states had such testing programs in place; today, virtually every state does. Legislators have begun applying the same logic to higher education: if we can define core abilities for high school students as a way of focusing attention on the central elements of an education, why can't we do the same for college students?

A third use of assessment is as a shorthand way of focusing on the "value added" by postsecondary education. Under this approach, students receive pre- and post-tests, and the gains in general education and skills are measured. The change in student performance is assumed to indicate the value added by the student's academic program.

The best illustration of this may be Northeast Missouri State University, which has employed such a system since 1974. The University's multi-faceted approach includes standardized tests for freshmen and sophomores, and major field examinations for graduating students (such as a Graduate Record Examination subject test, or occupational licensure examinations). The university also conducts attitudinal surveys of students and alumni.[8]

A fourth dimension of assessment in higher education refers to the use of standardized tests to examine either general or specialized knowledge. There are several testing instruments available for either type. The Undergraduate Assessment Program (UAP), developed by the Educational Testing Service (ETS), measures student knowledge and grasp of basic concepts in broad areas of the liberal arts. However, this program was largely discontinued in 1982.

The College Outcomes Measures Project of the American College Testing Program (ACT-COMP) can be used to assess general knowledge. This is a more application-oriented test than the UAP, and measures knowledge and skills against practical problems adults are likely to encounter. The test focuses on three subject areas (functioning within social institutions, using science and technology, and using the arts), and three process areas (communicating, problem solving, and clarifying values). More than 250 institutions now use this examination.

Fifth, some observers (especially state policy makers) see assessment as a way of making decisions about funding by rewarding institutions for student performance on established criteria. The leading example is undoubtedly Tennessee's Performance Funding Program. This effort emphasizes learning in general education, in the major field, and satisfaction with the educational experience. At the University of Tennessee, Knoxville, the Performance Funding Program has allowed some departments to relate student learning to curricular offerings and the University's budget process.

Finally, assessment sometimes refers to measuring changes in student attitudes and values. Higher education is justified in part by the "good neighbor benefits" it cultivates: open-mindedness, tolerance, interest in the community, and self-esteem. A number of survey instruments examine attitudinal development and growth, such as the American Council on Education/Cooperative Institutional Research Program and the College Board/NCHEMS Student Outcomes Information Service. Many institutions use their own instruments to measure change in this area.[10]

This brief review of the discrete items that are often found under the assessment umbrella illustrates the range of activities taking place, and demonstrates the importance of some precision in use of the word assessment. As originally designed and as adopted by the private sector, assessment is a valuable and powerful tool. But it is expensive and rarely what we mean when we speak of assessment in higher education.

What assessment appears to have become in higher education is a catch-all phrase that refers to a wide range of efforts to improve educational quality. This tendency to use one concept to refer to a handful of different (if related) things means that there are few shared meanings and little agreement about the nature, purpose, or content of appropriate public policies. Nonetheless, upgrading the educational quality of higher education—often in the name of assessment—will be a growing interest of state policy makers and an increasingly important challenge to educators in the next decade.

Assessment as Testing

The aspect of the assessment movement that has generated the most attention is student testing. There are three separate but related ways that states (and some institutions) are attempting to measure student performance through testing. The first tightens admissions standards to insure that students learn basic academic competencies in high school. In addition to testing, this approach often includes efforts to increase the number of academic courses required for college admission. A second approach more or less gives up on high schools and tests students at some point during their college career to insure specified levels of achievement have been reached. A final method imposes a graduation test as a way of guaranteeing that students meet at least minimum performance levels before receiving a college degree. Each of these approaches—testing to measure skills as part of the admissions process, to decide whether a student is sufficiently prepared to advance, or as a hurdle to graduation—merit some discussion.

Admission/Placement Testing. Standardized tests for students before they enroll in college have been an established part of the landscape for many years. Some institutions have simply responded to the interest in quality by raising admissions requirements on standardized tests—the Florida State Universities now require entering students to achieve a combined SAT score of 840 (Peebles, 1985). Nobody refers to such steps as assessment, nor does anyone really believe these actions will result in significant increases in educational quality at the college level.

Some states, however, have begun to test potential students more thoroughly. Florida, for example, requires all potential students at public institutions to take one of four approved standardized examinations. This serves several purposes: it permits a comparison among the colleges, provides a report card on secondary schools, and identifies students needing remedial assistance. Students who do not achieve a specified cutoff score on the test are admitted, but assigned to remedial courses. Because results from these four examinations are not easily comparable, the state is considering the possibility of requiring a single exam—either national or state developed—for all college students.[11]

Tests are also used to help make decisions about student placement and remediation. Perhaps the best known example is the New Jersey College Basic Skills Placement Test. The exam, developed in cooperation with the College Board and Educational Testing Service, consists of an essay and four multiple choice sections: elementary algebra, computation, reading comprehension, and sentence sense. Results are used for counseling and placement. The test is now administered at all the state's public colleges and at a number of private institutions that participate voluntarily.[12]

A variation on this approach comes from Ohio. Under the Early Testing Program administered by the Board of Regents, high school juniors take a version of the mathematics placement exam used by the state's public colleges and universities. Students are given information about their likely placement while they still have an additional year to take courses and address deficiencies. The program has resulted in increased mathematics enrollment among high school seniors, a higher level of mathematics readiness among college freshmen, and reduced enrollment in remedial courses. The state has recently implemented a similar program to assess the writing skills of high school students.[13]

Achievement Testing. In some cases, testing is used as a promotional gate to determine a student's readiness to move from one level of education to the next. One example of such a test can be found in "rising junior" examinations, so called because passage is required before a student is admitted to upper-class status (e.g., the junior class). The leading example of such an examination is Florida's College Level Academic Skills Test (CLAST). In August 1984, Florida required that all students in community colleges or state universities present passing scores on a state examination before receiving an associate's degree or enrolling in upper-division courses. The requirement has since been broadened to include all students who receive financial aid from the state, meaning that some students in private colleges are now tested as well.

The CLAST exam measures communication and computation skills including reading, writing (including an essay), and mathematical algorithms, concepts, generalizations, and problem solving. About 87 percent of the students taking the exam in August 1985 passed, but the state will raise the passing score in the next year, a move that may reduce the pass rate. All students receive score reports and interpretive guides, as well as information regarding performance on each of the tested areas.

Florida has supplemented the CLAST examination with curricular standards mandated by the state legislature. The so-called "Gordon Rule" named after its sponsor, State Senator Jack Gordon, requires all students to complete 12 semester hours of course work in English (including written work of at least 6,000 words in each three-hour course), and six semester hours of mathematics (at the level of college algebra or above).[14]

Only one other state (Georgia) currently mandates a statewide rising junior exam, but several others (including New Jersey and Texas) are considering such a test (*Change*, 1985). However, several individual institutions or public college systems have adopted their own version of a "rising junior" examination. The City University of New York uses the Freshman Skills Assessment Program to insure reading, writing, and mathematics proficiencies. The University of Arizona requires students to pass a writing proficiency examination near the mid-point of their undergraduate career. The University of Massachusetts at Boston requires undergraduates to pass a writing proficiency examination before they can take upper-division courses (Bennett, 1984).

Some states and institutions make students take examinations if they plan to enter certain areas of study. In recent years, several states have instituted a general education skills test for students seeking admission into teacher education programs as a way of insuring that only qualified students become teachers. Mississippi, for example, requires minimum scores on the ACT COMP examination. Other states have established a minimum score for prospective teachers on the Scholastic Aptitude Test. A recent survey by the American Association of Colleges of Teacher Education found that 64 percent of their membership now use some kind of test to screen candidates for admission to teacher education programs.[15]

Testing for Graduation. There can be a thin line between promotional gate testing and graduation testing. Florida's CLAST exam, for example, is clearly a graduation test for community college students and a gate for those in four year institutions. But beyond this, there are few examples of true graduation tests where students who do not pass the examination do not receive a degree. Despite the inroads state governments are making on the academic independence of colleges, they have been reluctant, so far, to impose graduation tests.

Perhaps the leading example of such an examination comes from Georgia. Beginning in 1973, the state required students to pass its "Regents Exam" in order to graduate. The two-hour test has a reading and essay section and is evaluated at state scoring centers. Although passage is required for graduation, students first take the exam as sophomores and retake it until they pass. In recent results, about 75 percent passed the reading section and 60 percent passed the writing part on the first try.[16]

Part of the difficulty in designing a graduation test for college students is the diversity of American higher education. The absence of a standard curriculum or an agreed upon central core of knowledge makes it difficult to develop a general-knowledge measure that would be suitable for all students across all institutions. Tests of basic skills—reading, writing, mathematics, etc.—may well insure an acceptable level of minimum competency for college students, but they will hardly suffice as the mark of an educated person.

Policy Considerations and Unsettled Issues

The extensive range of activities going forward under the assessment banner illustrates the widespread state and, to a lesser extent, institutional interest in insuring student achievement in higher education. The efforts so far appear to have been reasonably well designed. Still, there are reasons for concern. Much of what we refer to as assessment is really achievement testing by any other name, a much narrower, though important, activity. As well, the current activities raise a number of broader long-range questions that need to be addressed.

Some of the issues that should be of greatest concern to educators and policy makers alike are outlined in this section. The solutions to these issues are often obscure or difficult. Nonetheless, how they are answered will have an important bearing on the evolution of the drive toward improved quality.

What Is Quality? Any attempt to measure student outcomes quickly leads to questions about the goals of education; results cannot be assessed except in relation to the desired ends. And if the goal is quality, how do we define it? Some educators, such as former Ohio State University President Harold Enarson, claim that many efforts to measure quality are little more than "bush-league economics. It is zeal for quantification carried to its inherent and logical absurdity" (1983, p. 8). From this perspective, trying to specify and measure educational quality is likely to complicate the broader goals of learning, leaving students with only a cheap (but empirically verifiable) imitation.

Agreeing with this point does not mean all efforts are futile. Some efforts at assessment, such as Alverno College's comprehensive program, are rich and valuable tools. However, this approach will not work everywhere: it is expensive, time consuming, and requires a high degree of consensus about institutional goals. Moreover, there are enormous differences in scale involved. Alverno, with its 1,400 students, is a far more homogeneous place than Ohio State with an enrollment in excess of 50,000. At many large institutions, undergraduate education ranks, in truth, as the third or fourth priority and nobody is really in charge of it. In this environment, the incentives generally favor the status quo.

But these factors, while important, can easily become an excuse for not taking action. The question is less the size and structure of an institution than it is recognizing the growing public demands and acting upon them. There is, for example, nothing that precludes a university from establishing a general framework and guidelines and giving individual schools, colleges, or departments the responsibility for implementing appropriate steps.

The major barrier to taking action is that measuring educational achievement may well require more agreement about the ends and means of a higher education than exists at most institutions. It is possible to define a minimum level of information or skills that students should possess, sort of a least common denominator approach to college. But defining a general core of liberal learning and developing tools to insure that students are both broadly educated and deeply versed in a particular discipline is a far more complex task. State governments and coordinating agencies can do (and are doing) the former, but only institutions can do the latter. The most comfortable approach to defining quality may well be letting outside bodies do it, but this may cheapen public perceptions of higher education (it's a little hard to talk about higher learning when somebody is giving your students minimum competency tests) and erode institutional autonomy.

Achievement and Student Access. The growing interest in quality does not mean diminished support for expanding access to disadvantaged groups. Indeed, access as a policy objective is so widely accepted that no knowledgeable observer proposes anything but greater efforts in this direction. Nonetheless, there is concern that raising educational standards, at whatever level, will reduce minority enrollment in higher education. Indeed, the current emphasis on testing and measurement relies heavily on standardized instruments that have always been troublesome for minority students.

Reconciling equality and excellence has always been a difficult assignment and it will be no easier now. In fact, the challenges to be faced on the campus will be greater than ever before; colleges must simultaneously expand access to disadvantaged students and improve the quality of education they receive. This will require redoubling efforts to provide effective remediation both before and during the college experience. Such efforts will, of course, have implications for both staffing and funding. State governments are likely to be favorably disposed to the need for resources in this area; no state legislature will willingly accept a program designed to insure quality that fails large numbers of minority students. But remediation must now be seen as strictly temporary—the goal must be to bring students into the academic mainstream as quickly and efficiently as possible. Too often in the past remedial courses have become a substitute for meaningful and rigorous work.

The Cost of Quality. Raising academic standards will not be free. Even at the most basic level of adding an examination program, money is required to design and pretest the instruments, administer them, score and evaluate the results, and distribute scores to students and institutions. More elaborate assessment programs will involve greater costs. Related activities, such as remediation programs, will push the bill even higher.

But the resources required need not be excessive. New Jersey spends about $500,000 to have a contractor administer the state's Basic Skills Placement Test, and Florida spends a similar amount administering the CLAST program. Ohio spends $300,000 a year on the Early Testing Program, costs they believe are completely offset by the reduction in remedial education at the postsecondary level. At the institutional level, Northeast Missouri State University estimates annual costs of $60,000 (roughly $8.60 per student) for its comprehensive program.[17]

Even if the costs turn out to be greater than these illustrations, state governments have already indicated a willingness to spend more money on education. But, as the recent efforts to improve elementary and secondary education illustrate, there is an explicit *quid pro quo* involved. Higher funding for higher standards is possible. Higher funding without quality improvements is increasingly unlikely.

Making this even more likely is the growing competition for public sector resources. State efforts to improve precollegiate education will cost a great deal of money and, in some states, elementary school enrollments are increasing while postsecondary enrollments are stable or declining. This means that colleges, more than ever before, will be in direct competition with elementary schools and other social services for public funds. In this environment, clear, convincing evidence of higher quality might well allow institutions to make a stronger case for greater public support. Charles McClain, the president of Northeast Missouri State University, has repeatedly said that the positive results of his school's value-added program have made it easier to maintain support in the state legislature.

Legal Issues. Any assessment program that ties promotion or graduation to performance on standardized tests raises legal questions. While lawsuits aimed at blocking statewide or institutional testing programs at the postsecondary level remain comparatively rare, some have been filed. In Texas, for example, Federal Judge William W. Justice recently issued an injunction that forbids the state from requiring teacher education students to pass a Pre-professional Skills Test.[18] How this and similar cases will be resolved is unclear, but the extensive record of such suits at the elementary and secondary level indicates that caution, and careful design, will be essential. Mingle (1984) suggests that, at a minimum, three considerations should be kept in mind: Has adequate notice of the program been given? Are the test materials racially or culturally biased? Does the test reflect the material taught? The last issue may be the most important; any measurement instruments must be sufficiently related to curricular offerings to withstand judicial scrutiny.

Is Assessment Tied to Funding? Funding for public colleges and universities historically has been based on enrollments and the kinds of programs offered rather than how well students were educated. In recent years, enrollment-based funding encouraged institutional growth and an expansion of student access. At the same time, state governments were often hesitant to use performance criteria in the budget process because it raised difficult questions about definitions of quality and measurement of performance. Institutions were no more anxious to rely on performance standards than were state governments. Now, as educational quality becomes an important policy focus for state governments, there are suggestions that funding formulas should also be modified.

Several models may be used. One is performance-based budgeting that rewards institutions for meeting specified goals. Tennessee has such a system; it lets institutions supplement their core budget by demonstrating progress toward agreed upon measures of improved quality. A second approach is to establish and announce performance goals and outcome measures that will serve as a benchmark for evaluating institutional efforts. This approach does not tie funding directly to results, but it does provide a target that is likely to be considered in making budgetary decisions. Florida and several other states have expressed interest in this approach.[19]

Yet another way to encourage improvement efforts is to create a separate source of money that permits institutions to request money for quality enhancing projects. While such an approach does not relate quality improvements to state funding, it does have considerable appeal. The approach is popular with colleges since it permits them to decide when (and if) to undertake projects and allows a clear focus on local needs and interests. From the state's perspective, this can pave the way for "joint-ownership" of the effort by requiring cost-sharing and insures a favorable reception at the institution. The weakness is that support for separate funding is hard to maintain (or expand) and it may be difficult to make specific projects institution-wide priorities.[20]

Institutional Autonomy and State Authority—The drive toward higher standards in postsecondary education may jeopardize the American tradition of institution-based quality control. One educator has warned: "If American higher education is to forestall the imposition of a state system of examinations, it will have to improve its own forms of quality control. . . . If the academy does not strengthen these controls of its own volition, it may find government moving to do so in ways that jeopardize the core of the enterprise" (O'Neill, 1983, p. 78).

If the states take the lead, they will probably treat all institutions in a very similar, if not identical, fashion. Such an approach may undercut institutional autonomy, increase the homogenization of higher education, and stifle innovation. Should this occur, the diversity that we prize, and that the rest of the world admires, will be seriously undermined. Most of the state-level programs enacted so far have been carefully designed, but future initiatives may turn to standardized measures that can be administered cheaply and interpreted easily, perhaps even offering a single number as the current level of quality in individual colleges. Americans hunger for such information. Witness, for example, the reliance on SAT scores as a benchmark of secondary school quality, despite arguments by educators that the test is a poor instrument for such purposes. Imagine how college officials would react if the nation's GRE scores were mandated and released each year amid such media attention and public comment.

An additional danger in this regard harkens back to the previous policy issue—whether such scores are used to make budgetary decisions. If institutional funding is tied to results on state measurement instruments, faculty may feel pressured to teach to the test, especially if they in turn are evaluated on students' performance. There are some suggestions that "teaching to the test" already takes place in states where such programs exist (Rentz, 1979). While this insures that students have a basic floor of knowledge, it also diminishes institutional flexibility and autonomy.

Summing Up: It's Here to Stay

Concerns about what, if anything, colleges and universities teach their students are not new. Harvard's legendary president Charles Eliot, who virtually eliminated required courses for undergraduates, was once asked why Harvard was such a great storehouse of knowledge. "In all likelihood," he allegedly replied, "it is because the freshmen bring us so much, and the seniors take away so little." Throughout the long history of American higher education, we have experienced regular periods of concern that graduates were taking away too little knowledge from their college experience. We are now in another such era, and the move to assess student achievement flows from it.

The drive to insure quality raises a host of troubling issues, ones that go to the heart of the college experience and the relationship between higher education and the many publics it serves. Some in higher education hope that this is nothing more than a passing fancy. Colleges and universities are very conservative institutions in which change comes slowly, if at all. Those who advocate large-scale assessment would appear to want colleges to plunge off into a brave new world with few road maps. Asking colleges to do something they don't want to, that is only loosely defined, and that threatens to upset existing arrangements, has all the makings of a fad or a disaster. In either case, it should be avoided.

Assessment is not likely to be a fad. One reason that the standards issue will not go away is easily overlooked by educators. State governments, once the whipping posts of American politics, are more competent and professional than ever before. Constitutional modernization and administrative reform have transformed state capitals. State governments now ask more and better questions, have more information and assistance available to them, and are much more visible and active actors than they were twenty years ago. Legislatures and governors are increasingly asking what the state is getting for its money. The capacity to ask tough questions and the willingness to act means that colleges and universities can soon expect (and in some cases are already getting) the same sort of attention that has been given the public schools (Doyle and Hartle, 1985).

Most of the scrutiny in the future will be on public two-year and four-year colleges. Too often we use the leading research universities or selective colleges as the reference point in discussions about higher education. In reality, these institutions probably enroll less than 15 percent of the nation's students. But the mission of these schools has changed little in the last two decades and the competition for admission by students offers some assurance of quality. Community colleges and state colleges, however, serve all comers, and the mission of both types of institutions have grown more complicated (and obscure) in recent years. Many state legislatures regard these schools as directionless and mediocre.

This does not mean that private colleges occupy a completely safe harbor. Some states provide direct subsidies to their private institutions, and many others provide indirect assistance. Most states regulate at least some aspect of private higher education within their borders. If public funds support it, public regulations can follow, as Florida's expansion of the CLAST program to private college students receiving financial aid illustrates. Moreover, some private colleges are already desperate for students and willing to take anyone as a way of filling classrooms. State governments know that in such an environment, quality is too often a secondary consideration.

In short, concern with the outcomes of higher education and student achievement is likely to become an increasingly prominent part of the policy landscape. Higher education has two choices in this regard. It can wait, watch, and see how developments evolve. In the meantime, more states are likely to take action. Alternatively, colleges can take a leadership role and implement programs that meet the public interest while preserving institutional autonomy.

The latter course will require enormous leadership at the campus level. Unfortunately, the incentives often work against academic leadership by college administrators. One recent study of college presidents found that few of those surveyed described themselves as playing a major role in academic affairs (Kerr, 1984). This does not mean that college presidents can do it alone. Only by involving the entire college administration and staff is there a reasonable chance of success. In Education Secretary Bennett's words:

> Revitalizing an educational institution is not easy. Usually it requires uncommon courage and discernment on the part of a few and a shared vision of what can and ought to be on the part of many (1984, p. 25).

Most state legislatures would prefer to see colleges and universities take the lead in this area. Legislators recognize the complexity of the issues involved, and the political rewards involved are not great. Self-regulation is a popular public policy tool these days if it serves the public interest in a clear and appropriate fashion. Strong steps toward institutional renewal will be well received in state capitals. But legislators will not be satisfied with bland assurances of quality, or meaningless indicators.

Whether higher education institutions can marshall the leadership, energy, and creativity to meet the quality challenge by themselves remains to be seen. But one thing is clear: the issue will not quickly fade away.

Notes

1. See, for example, the report of the Southern Regional Education Board. *Access to Quality Undergraduate Education*. Reprinted in *The Chronicle of Higher Education*, July 3, 1985, pp. 9–12.

2. Clifford Adelman. "The Standardized Test Scores of College Graduates, 1964-1982." Washington, DC: National Institute of Education, 1985. In June 1985, *The National Law Journal* reported that scores on the 1984 bar exam were the lowest since 1979. More specifically, the percentage of applicants passing the exam fell in 32 of the 54 jurisdictions administering a bar exam in July 1984. Bar examiners attributed the result "to a variety of reasons, ranging from an increase in applicants who are incompetent or cocky—or both—to a decrease in the quality of cram courses. See David Kaplan "Bar Exam: The Rites of Passage Are Getting Tougher." *The National Law Journal*, Monday, June 3, 1985, p. 1 ff.

3. See, for example, Scott Jaschik, "States Questioning Role of Colleges in Remedial Study." *The Chronicle of Higher Education*, September 11, 1985; Scott Jaschik, "States Are Urging Community Colleges to Review Programs, Justify Expenses." *The Chronicle of Higher Education*, October 16, 1985.

4. Ernest Boyer. *College: The Undergraduate Experience in America* (New York: Harper and Row, forthcoming) cited in Andrew Hacker, "The Decline of Higher Learning." *The New York Review of Books*, February 13, 1986, p. 36.

5. See for example, Chester E. Finn, Jr. "Trying Higher Education: An Eight Count Indictment." *Change Magazine*, May/June 1984, 29ff; Alan Bloom, "Our Listless Universities." *National Review*, December 10, 1982, pp. 1537–1548; Timothy Noah, "Highbrow Robbery: The Colleges Call It Tuition, We Call It Plunder." *The Washington Monthly*. July/August 1983, pp. 16–25; Andrew Hacker, "The Decline of Higher Learning."

6. Scarvia B. Anderson, Samuel Ball, Richard T. Murphy and Associates. *The Encyclopedia of Educational Evaluation* (San Francisco: Jossey-Bass, 1975), pp. 26–29.

7. For a summary of the Alverno program see the American Association for Higher Education *Bulletin*, Volume 36, no. 6, February 1984, articles by Russell Edgerton and Marcia Mentkowski and Austin Doherty.

8. For an excellent overview of the Northeast Missouri program, see Charles J. McClain and Darrell W. Krueger, "Using Outcomes Assessment: A Case Study in Institutional Change." in P.T. Ewell, ed. *Assessing Educational Outcomes* (San Francisco: Jossey-Bass, 1985), pp. 33–46. American Association of State Colleges and Universities, *In Pursuit of Degrees with Integrity: A Value Added Approach to Undergraduate Assessment* (Washington, DC: AASCU, 1984). For a shorter summary see Clifford Adelman, *Starting with Students*: Washington, DC: National Institute of Education in cooperation with the American Association for Higher Education, 1985.

9. Peter T. Ewell, "Levers for Change: The Role of State Government in Improving the Quality of Postsecondary Education." National Center for Higher Education Management Systems. July 1985. John Folger, ed. *Financial Incentives for Academic Quality* (San Francisco: Jossey-Bass, December 1984).

10. For a discussion of "good neighbor benefits" see The Carnegie Commission on Higher Education, *Higher Education: Who Pays? Who Benefits? Who Should Pay?* (New York: McGraw-Hill, June 1973), 81ff. Alexander Astin, "Why Not Try Some New Ways of Measuring Quality?" *Educational Record*. Spring 1982, pp. 10–15.

11. James R. Mingle, "Measuring the Educational Achievement of Undergraduates: State and National Developments" (Denver, Colorado: State Higher Education Executive Officers, 1985). photocopy.

12. For a summary of the New Jersey College Basic Skills Placement Test see Alice J. Irby, "Statewide Assessment of Basic Skills" in *Current Issues in Higher Education 1979: Assessment* (Washington, DC: American Association for Higher Education, 1979), pp. 11–16.

13. For a summary of Ohio's efforts see Board of Regents, "Articulation Efforts Proving Successful in Ohio." *Regents Review* (February 1986); Ohio Board of Regents, "Early English Composition Assessment Program Description," 1986, photocopy; and Bert K. Waits, "The Ohio Early College Mathematics Placement Testing Program for High School Juniors" *Final Report 1983–85 Biennium*, July 1985, photocopy.

14. See Mingle, "Measuring the Educational Achievement of Undergraduates . . ." and John Bennett, et al. "Academic Progression Tests for Undergraduates: Recent Developments" *Educational Record*, Winter 1984, pp. 44–48.

15. *Ibid*. See also American Association of Colleges for Teacher Education, "1984 Report to the Profession: Data Show Innovation, Change" *AACTE Briefs* (Washington, DC: July 1984).

16. *Ibid*., see also R. Roberts Rentz, "Testing and the College Degree," *New Directions for Testing and Measurement* (San Francisco: Jossey-Bass, 1979), pp. 71–77.

17. Estimates of program costs furnished by officials at Florida Board of Regents, Educational Testing Service, Ohio Board of Regents, and Northeast Missouri State University. In all cases, the estimates refer only to direct costs and do not include the full range of indirect charges that such programs involve.

18. See "Texas Appeals Injunction of Teacher Education Test," *The Chronicle of Higher Education*, September 18, 1985, p. 16.

19. Trudy W. Banta and Homer S. Fisher, "Performance Funding: Tennessee's Experiment," in John Folger, ed. *Financial Incentives for Academic Quality* (San Francisco: Jossey-Bass, December 1984), pp. 29–42. E. Grady Bogue and Wayne Brown, "Performance Incentives for State Colleges," *Harvard Business Review*, Vol. 60, no. 6, November-December 1982, pp. 123–128. See also "Editor's Notes" in John Folger, ed. *Financial Incentives for Academic Quality* (San Francisco: Jossey-Bass, December 1984), pp. 1–13.

---

20. Editor's Notes'' in John Folger, ed. *Financial Incentives for Academic Quality*, p. 10. Folger notes that this approach might include a competitive grant program with peer review of proposals. He suggests that the federal government's Fund for the Improvement of Postsecondary Education may provide a model for such an effort.

References

Adelman, Clifford. ''The Major Seventh: Standards as a Leading Tone in Higher Education'' in J. R. Warren (Ed.) *Meeting the New Demand for Standards*. San Francisco: Jossey-Bass, 1983, pp. 39–54.

Association of American Colleges. *Integrity in the College Curriculum*. Washington, DC: Association of American Colleges, 1985.

Bennett, John, et al. ''Academic Progression Tests for Undergraduates: Recent Developments.'' *Educational Record*, Winter 1984, pp. 44–48.

Bennett, William J. *To Reclaim a Legacy: A Report on the Humanities in Higher Education*. Washington, DC: National Endowment for the Humanities, 1984.

Change Magazine, ''State Initiatives: An Update,'' November/December 1985, pp. 16–17.

Doyle, Denis P. and Terry W. Hartle. ''The States are Leading as Washington Wallows.'' *The Washington Post Outlook*. September 8, 1985.

Enarson, Harold L. ''Quality—Indefinable But Not Unattainable.'' *Educational Record*, Winter 1983, pp. 7–9.

Kerr, Clark. *Presidents Make a Difference: Strengthening Leadership in Colleges and Universities*. Report of the Commission on Strengthening Presidential Leadership (Washington, DC: Association of Governing Boards, 1984).

MacKinnon, Donald W. ''An Overview of Assessment Centers.'' Greensboro, N.C.: Center for Creative Leadership, 1975. photocopy.

O'Neill, Joseph P. ''Examinations and Quality Control.'' in J.B. Warren, ed. *Meeting the New Demand for Standards*. San Francisco: Jossey-Bass, September 1983.

Peebles, Lynn. ''Higher Standards for College Entry Viewed by States.'' *The Chronicle of Higher Education*, November 6, 1985.

Rentz, Robert R. ''Testing and the College Degree.'' *New Directions for Testing and Measurement*. San Francisco: Jossey-Bass, 1979.

Southern Regional Education Board. Commission for Educational Quality. *Access to Quality Undergraduate Education*. Reprinted in *The Chronicle of Higher Education*. July 3, 1985, pp. 9–12.

Study Group on the Conditions of Excellence in American Higher Education. *Involvement in Learning: Realizing the Potential of American Higher Education*. Washington, DC: National Institute of Education, 1984.

U.S. Department of Education. ''Recent Initiatives Reported by States and the District of Columbia.'' Washington, DC: USDE, 1984. photocopy.

Assessing Outcomes in Higher Education

by John Harris

This paper is intended to offer practical advice on assessment of educational outcomes to a chief academic officer. The use of the first and second person is intended to convey the directness of a consultant's report.

I have assumed that, as the chief academic officer, you are trying to get started in outcomes assessment. Therefore this report is composed of suggestions of critical issues to consider, organizations that can help, and what assessment approaches and instruments you might use.

I. Goals

You can compare your students to other students nationally on standardized tests without having definite educational goals, stated expectancies, or outcomes. But without such goals, you can't be sure the tests reflect your curriculum. You and your colleagues may also be interested in how your students change in terms of their beliefs, interests, attitudes, values, and behaviors. There are various commercially available inventories to reflect these things. Yet again, without relatively clear student development goals, you won't know how to select the inventories that fit your institution.

Responsibilities

If you are without clear goals for student academic achievement and personal development, I hope you will seriously consider developing some. If you decide to develop student achievement goals, the first step is to decide who will be responsible for their development. While respective departments may propose goals, they should be reviewed, possibly modified, and eventually owned by a committee or council representative of the whole institution.

Both department and larger institutional committees will be faced with the dilemma of "specificity" versus "consensus." The more specific your goals, the better it is for instructional clarity and for the conduct of assessment. Yet the greater the specificity, the greater the difficulty in reaching campus or departmental consensus. There is no easy answer to this dilemma. Realize from the beginning that the articulation of specific educational goals by faculty consensus will require a great deal of patience and diplomacy.

Specificity

How specific should goals for general education and majors be? They have to be specific enough so that two faculty members independently writing test items or designing exercises or projects to reflect them, come up with roughly the same type of items, exercises, or projects. Basically, goals ought to describe observable performances or products. The verbs in goal statements tell one a great deal. The better goal statements use verbs such as "paraphrase," "compute," "describe," and "construct." The poorer ones use more general verbs such as "appreciate" and "understand."

A Beginning

A productive strategy for developing goals begins by asking the following basic questions:

1. What do you implicitly expect of all students and graduates in terms of knowledge, skill, attitude, and behavior?
2. What achievements do you implicitly expect of graduates in each major field?
3. What profiles of your alumni do you have, or can you develop, in terms of such achievements as career accomplishments, lifestyles, citizenship activities, and aesthetic and intellectual involvements?

Item #1 can be answered by identifying common proficiencies required in the assignments and examinations of the general education courses. Similarly, Item #2 can be answered by identifying the knowledge and skills usually reflected in the examinations and assignments in major courses. You might ask an expert in test development currently in your faculty or administrative staff to develop a simple two-dimensional table for a "content" and "mental process" analysis of test items. Make sure the test expert develops a form that is understandable and useful to his or her colleagues. With his or her help, faculty in the respective disciplines can sort their test items by level of thought process and area of content.

The personal development goals related to Item #3 are usually difficult to define. While they should reflect the values of the institution and its constituencies, our increasing intra-institutional pluralism makes agreement on specific personal goals very difficult. Nevertheless, most campuses will agree to such goals as sensitivity to, and awareness of, civic responsibilities, preference for democracy or autocracy, and vocational success. These developmental goals often blend with general or liberal educational goals.

Outsiders

Academics, as any professionals, need the perspective of outsiders. That is, when they are developing general education goals, they need to think seriously of what the larger world expects of college graduates. You might include some people from outside of your institution in the process of developing and reviewing goals. Try to find outsiders who are not intimidated by the parlance, rites-of-passage, or bureaucracy of academia. For example, business executives, foresters, elementary school teachers, artists, and others not employed in higher education can have a keen sense of the common skills needed by college graduates.

II. Given the Goals, Why Assess?

In my judgment, there are two primary reasons behind the current emphasis on assessment:

1. Concern that college graduates have the abilities that their degrees are supposed to certify.
2. Need for a more direct way to determine the effectiveness of instruction.

In contrast to the manufacturing paradigm, higher education is without direct indicators of quality assurance. Most of its indicators of effectiveness have to do with the "richness" of its processes, i.e., credentials of faculty, classroom, and laboratory facilities, work loads of faculty, instructional technology resources, etc. In contrast, the interest in outcomes assessment is intended to move us toward "tasting the pudding," in addition to checking on the cook and the ingredients.

The American academy is quite vulnerable on the issue of quality assurance. O'Neill (1983) argues forcefully that the integrity crisis is rooted in the arrangement under which the same individual who instructs a student also tests and certifies his learning. Wang (1975) nipped at the academy's heels for "bundling" its services of imparting information, accreditation, coercion (structure), and club membership. He suggested that if colleges and universities were commercial institutions, they would be in violation of the Sherman Antitrust Act for "bundling" these services.

The point is this: unlike British or European institutions, our certification of student achievement is done by the same person who teaches the student. Related to this linkage, we have also chosen to report educational progress in proxy time measures (credit hours), rather than units of achievement. As a result, the system is very vulnerable to compromise of standards by grade inflation and consequent devaluation of degrees. Because our current indicators of educational quality depend heavily on "richness of treatment" and "time," we are limited in controlling quality in terms of results. Without outcomes assessment, we appear to believe that the more it costs, and the longer it takes, the better it is. The first step toward change is to make a separation between the "means" (instruction) and the "ends" (achievement outcomes).

An Ideal Goal

As one primarily interested in the systematic improvement of instruction, I believe some "unbundling" of testing from instruction would be helpful. To be improved, instruction in any subject must be judged in terms of its *effects* (how much and how well have students learned), *costs* (in terms of effort, time, and resources compared to learning), and *acceptance* (students' identification with particular instructional approaches).[1] By separating assessment of student achievement from instruction, we are more likely to compare modes of instruction in terms of their effectiveness, efficiency, and acceptance.

The self-contained course and our time-based method of accounting for educational attainment in American higher education work against such separation. There are inherent difficulties in evaluating instruction where credit for a degree is counted directly in time units (credit hours) and only indirectly in amount learned. Furthermore, with instructional goals and testing patterns being almost as different as the teachers in different courses, there is no common standard by which to evaluate instruction.

An increase in *external* assessments will likely continue until there is some operational separation between instruction and assessment within our institutions. There are least two steps faculty and administrators might take to connect instruction as a means, with assessed achievement as an end:

1. Institute or reinstitute the senior comprehensive, as suggested earlier. Arrange for the faculty member who directs and instructs in the comprehensive to present his or her students to a panel of examiners. Perhaps the panel of examiners could be composed of other faculty members from on- or off-campus. In some areas, involve off-campus, practicing professionals where the major leads directly to a professional or technical vocation.
2. Require common, comprehensive examinations or papers for the basic general education courses expected of all students. Ask the faculty teaching those courses to work together, and possibly with a test development specialist, to construct comprehensive examinations or assessment procedures. If there are essay responses or student performances or products that have to be graded subjectively, ask the faculty to develop a system for at least two graders to independently assess each student's work.

Senior comprehensives with multiple evaluators and common assessment of general education skills and knowledge will inevitably serve as strong catalysts for instructional improvement. Furthermore, both of these are consistent with academic traditions with which most faculty can identify.

In addition to seniors doing major papers and projects in their comprehensives, you may occasionally choose to administer appropriate nationally standardized tests to seniors in each major field. Despite my emphasis on the senior paper or project, faculties need to know how their students compare nationally.

III. Test Selection

Before considering some commercially-available examinations, you may find a suggested technique of analyzing tests helpful. Specific student learning goals for general education and majors become very helpful at this point. Morris and Fitz-Gibbon (1978, pp. 47-68) have developed a procedure that a faculty committee could use to determine if a given test fits particular programs, including how to "refine and organize program objectives" and how to "estimate the relative match of the test items to program objectives."

Using this procedure, your faculty can determine if a particular test fits a particular program. On the other hand, you might ask someone on your campus with competence in test development to construct a system of comparing test items to program content. For example, they might analyze an American history examination by placing items in the appropriate cells of a table similar to the one below.

Early American History

		Content: Historical Periods		
		Exploration	Colonization	Revolution
Process: Levels of Thought	Apply Facts and Concepts			
	Comprehend Concepts			
	Recall Facts			

Through a table with course or program "content" on one dimension and mental "process" on the other, individual test items may be placed in the appropriate cells. Once all items are distributed in such a table, it will be easier for a faculty to determine if a given test's items are congruent with the objectives of a course or program. For more information on how to construct such a table, see Scannell's and Tracy's *Testing and Measurement in the Classroom* (pp. 49-69).

Selection vs. Criterion-Referenced Tests

The United States has led the world in the production and use of standardized, objective tests for selection purposes. The focus of *selection* tests has not been to compare a student's performance to an absolute standard of knowledge or skill, but to the performance of others. The scoring and scaling methods of selection tests are intended to maximize individual differences for purposes of comparison.

In contrast, the historic intent of educational tests is to determine how much of a body of knowledge one knows, or how skillful one is as compared to some pre-set standard. In more recent years, psychologists have referred to these as criterion-referenced tests.

The two types of tests are developed differently. Ideally, the selection test excludes items that are very frequently answered correctly or incorrectly. The ideal selection test item is one that 50 percent of the students answer correctly. Let's assume a given item accurately reflects a critical skill, but no one answers it correctly; following the selection test approach, it would be deleted. Conversely, if everyone answered it correctly, it would still be deleted.

Now let's say a teacher developed a very effective instructional program in general biology and the students were all able and motivated. Further assume the teacher taught well and the students studied effectively so that all of them answered every item on the final examination correctly. Using the selection test, item-analysis approach, the test is at fault because it does not discriminate among the students. The instructor should continue developing items until significant percentages of students miss each item. By the selection test standard, 50 percent of the class should miss each item. By the time our hypothetical biology instructor using the selection test approach has reached this point, he or she is assessing differences in individual native intelligence more than mastery of the specific content of what has been taught or learned.

The selection test approach works well when the purpose is to spread individuals over a continuum. But it is awkward, to say the least, when the purpose is to certify a level of competence. It is also questionable when the purpose is to assess the impact of instruction on a group of students. Its difficulty lies in its emphasis on differences between an instructed group and an uninstructed one. The selection test approach so strongly emphasizes variation in individual ability that the differences of individuals' scores within the instructed group will often be greater than the distinctions between instructed and uninstructed groups. The same is true for differences among individuals within groups that have been instructed in various manners, i.e. lecture, discussion, or structured independent study.[2]

Nevertheless, the commercially available achievement tests you will come across have been built, for the most part, on the selection model. In practical terms, this means you will be working against the odds to show significant gains in scores over time if you use such instruments. You will encounter the same problem in attempting to demonstrate the differential impact of various instructional approaches. On the other hand, if your primary purpose is to compare the performance of your institution's average student to the performance of students in similar programs nationally, then nationally normed, standardized tests built on the selection model can be helpful.

If you use a usual standardized test to compare possible gains in knowledge or skills, or to compare different instructional approaches, first ask the test publisher if the scores can be interpreted in a criterion-referenced way. If not, be prepared for the differences to be insignificant and do *not* assume that the lack of significant differences is completely attributable to ineffective instruction.

Basic Skills

Since this advice is on how to assess outcomes to improve instruction, some assessments are suggested for use at the input stage of general education as well as at the outcome point. Input assessment helps one focus instructional time and resources on deficiencies of individual students; outcome assessment provides feedback on the effectiveness of instruction once it has occurred.

A useful summary of the skills and knowledge needed by entering college students is *Academic Preparation for College: What Students Need to Know and Be Able to Do*. (New York: The College Board, 1983).

This publication describes the basic academic competencies expected of entering freshmen, as well as expected mastery of content in the basic academic subjects of English, the arts, mathematics, science, social studies, and foreign language. To match these statements of expectations, the College Board's Multiple Assessment Programs and Services (MAPS) provides a comprehensive diagnostic assessment for advising and placement. It includes:

1. Descriptive Measures of Students
2. Vocationally Oriented Measures of Interests and Special Aptitudes
3. Measures of Basic Reading, Writing, and Mathematical Skills
4. Measures of General Academic Potential (SAT)
5. Measures of the Ability to do Academic Work on an Introductory College Level in English, mathematics, natural sciences, social sciences, and foreign language and literature.

For complete information on both *Academic Preparation for College* and MAPS, write or call:

The College Board
45 Columbus Avenue
New York, NY 10023
(212) 713-8000

Components of MAPS are being used in various configurations in Tennessee, Florida, New Jersey, and California. In Tennessee, for example, the State University and Community College System has developed a comprehensive screening and placement system using MAPS tests. Any entering student with an ACT composite below 16 will be tested with MAPS tests. Given the student's MAPS performance, he or she will be placed in certain remedial or developmental courses. For more information on this screening and placement procedure, contact:

The State University and Community
College System of Tennessee
1161 Murfreesboro Road
Nashville, TN 37217
(615) 741-4821

In addition to these test batteries specifically designed to assess basic collegiate skills, there are other tests of prior achievement you might use to assess both general knowledge and basic skills of incoming students. References you could use in search for such tests are listed and described in Morris and Fitz-Gibbon (pp. 39-44). This list will also be helpful in considering tests to assess outcomes of general education and major fields of study.

Two more recent references that will be very helpful are:

James V. Mitchell, Jr., Ed. *Tests in Print III: An Index to Tests, Test Reviews, and the Literature on Specific Tests*. The Buros Institute of Mental Measurements. Lincoln: The University of Nebraska Press, 1983.

Richard C. Sweetland and Daniel J. Keyser. *Tests: A Comprehensive Reference for Assessments in Psychology, Education, and Business*. Kansas City: Test Corporation of America, 1983.

General Education

There are a few tests to assess outcomes of general education. As they are described below, be reminded of the importance of comparing these tests with the goals of your particular general education program.

One approach to assessing general education is a second administration of the ACT at the end of the sophomore year. Northeast Missouri State University readministers the ACT to about one-half of its sophomores. This allows for comparison of entering freshman and rising junior average scores on each of the four parts of the ACT.

Perhaps the most widely used general education battery is ACT's College Outcomes Measures Program's (COMP) Assessment. More than 250 colleges and universities have used COMP. There are three options in COMP:

1. The COMP Composite Examination covers three "process" and three "content" areas. The process areas are oral and written communication, problem solving, and values clarification. The three content areas are "functioning within social institutions," "using science and technology," and "using the arts." The examination includes multiple choice response questions, questions requiring brief written responses, exercises of writing letters and memos, and exercises requiring brief oral speeches. About four hours are required for a student to complete the entire examination. The evaluation of written and oral responses takes about 50 minutes per student tested by a four-person faculty evaluation team. The examination is modular so that you can select the parts you wish to use.[3]
2. The COMP Objective Test covers the same "process" and "content" areas as the Composite Examination, except the "communication" area is not divided into oral and written sections. The format consists entirely of four-option multiple choice questions requiring no faculty evaluation of responses. This test takes about two hours of a student's time to complete.
3. With the COMP Activity Inventory, students report activities and perceptions in the same three process and content areas assessed on the Composite and Objective Examinations. This inventory is not timed but, according to Forrest and Steele, students usually take about 90 minutes to complete it. The intended purpose of the Activity Inventory is to obtain a report from students or alumni of what uses they make of their general education.

The COMP Activity Inventory is a simulated version of what a friend of mine once described as the "candy store test." By that he meant you can best determine the impact of general education by observing what students do in free-choice situations. That is, what actual use would a graduate make of his or her general education literature courses in selecting a novel in a large bookstore? The Activity Inventory simulates situations to elicit the effects of general education upon: communicating about social institutions; solving social problems; clarifying social values; communicating about science and technology; solving scientific and technological problems; clarifying scientific and technological values; communicating about the arts; solving artistic problems; and clarifying artistic values.

The Activity Inventory asks the respondent to indicate what he or she usually, rather than ideally, does. In judging the results of this inventory, remember that typical, or usual, behavior is greatly influenced by immediate circumstances and pressures, and only indirectly by the perspective or insights of previous formal instruction.

If the items constituting the COMP options reflect the goals of your general education program, you might use them to compare your students to a national sample. I favor assessments that involve faculty in evaluating student responses as the Composite Examination does. If faculty are not involved, they may dismiss the assessment results. Furthermore, they will miss the face-to-face specifics of the students' responses, which encourage them where the students do well and provide them with specific knowledge of deficiencies where the students perform poorly. As with most proposed changes in higher education, faculty must be involved in order to be committed. However, I realize how much student and faculty time the Composite Examination involves, especially when such testing and grading are done outside the normal process of classroom testing. It is, thus, not surprising that institutions are using the Objective Examination.

The Educational Testing Service (ETS) at one time offered the Undergraduate Assessment Program (UAP). The UAP tests were derived from Graduate Record Examination (GRE) Subject Tests. Out of the original UAP, three general education area tests and one major field test, survive. The available General Education Area Tests are Humanities, Social Science, and Natural Science; each one is 60 minutes. The one major field test is a general test of business requiring two hours. ETS will loan these tests to an institution for a year. The institution must score its own answer sheets. Obviously, without ETS scoring there are no *current* national norms. If you wish more information, write or call:

ETS College and University Programs
Educational Testing Service
Princeton, NJ 08541
(609) 734-1162

ETS also offers General Examinations in English Composition, Mathematics, Humanities, Natural Sciences, and Social Sciences and History as part of the College-Level Examination Program (CLEP). This program was structured for individual high school students to take the examinations at ETS testing centers for college credit. These General Examinations have current national norms and would, therefore, allow you to compare your students against wider groups. For more information, contact:

College-Level Examination Program
The College Board
45 Columbus Avenue
New York, NY 10023
(212) 713-8000

By Florida Department of Education rules and state statue, every community college and state university student in Florida has to take and pass all four tests of the College Level Academic Skills Project (CLASP). Every community college student must take it to receive an A.A., and all state university students must take it to be admitted to upper-division status. CLASP assesses the communications skills of reading, listening, writing, and speaking. In mathematics, it assesses competence in algorithms, concepts, generalizations, and problem solving. This test battery was developed by faculty from the Florida community colleges and state universities. It is a secure battery, not for use outside its designated testing centers, and for Florida students only. Nevertheless, you may wish to have your faculty review its content and techniques of development and administration. To do so, ask for the *CLASP Technical Report 1982-83* and *CLASP Test Administration Plan 1984-85*. Write or call:

College-Level Academic Skills Project
Department of Education
State of Florida
Tallahassee, FL 32301
(904) 488-0325

The New Jersey Board of Higher Education has developed The New Jersey College Basic Skills Placement Test Program. It includes an expository essay and multiple choice questions on "reading comprehension," "sentence sense," "math computation," and "elementary algebra." This test program is administered to all students coming into public New Jersey colleges and universities, as well as eleven private New Jersey colleges. If you are interested in this test program, contact:

New Jersey State Board of Higher Education
225 W. State Street
Trenton, NJ 08625
(609) 292-4310

I have been struck by how much attention is being given to writing in state and system-level assessment. The California State University System, the Florida Department of Education, The University System of Georgia, and the New Jersey Board of Higher Education all require a demonstration of writing proficiency of college students either at entrance, at the rising junior level, or before exit. This confirms the general impression that the only common component of general education left within and among many institutions is a required course in composition.

The California State University System's Graduation Writing Assessment Requirement (GWAR) is implemented differently on each of the nineteen campuses in the CSU system. All upper-division and graduate students must demonstrate writing proficiency. Each campus reports how it certifies writing ability and the number of students who pass. Some campuses require students to take designated upper-division or graduate courses requiring a large amount of writing. Others allow students to demonstrate proficiency on a writing test. Your faculty may want to review some of the tests developed on different campuses. For more information contact:

Office of the Chancellor
The California State University
400 Golden Shore
Post Office Box 1590
Long Beach, CA 90801-1590
(213) 590-5480

The Regents' Testing Program of the University System of Georgia also requires students to produce acceptable essays. All rising juniors in all state community colleges, four-year colleges, and universities must take and pass the Reading and Essay Tests before they can graduate. The Reading Test consists of ten reading passages, with five to eight questions on each, that test comprehension in terms of vocabulary, literal comprehension, inferential comprehension, and analysis. The reading passages are selected from materials college graduates should understand. It is a one-hour test of 60 items.

The Georgia essay test, like those in New Jersey and the CSU system, uses multiple faculty evaluators (who are not directly involved in teaching the students) with a very consistent scoring procedure. There are several advantages in having these faculty judge students' work. First, this approach forces faculty to look directly at what students can do. Second, by having to explain their judgments to a second or third reader, faculty begin to develop a collective sense of what they expect. So, if you are primarily interested in outcomes assessment serving as a catalyst for instructional improvement, I suggest that you look for reliable ways to involve your faculty in directly evaluating students' performances and products. For information about the way the Georgia Regents Testing Program does this, contact:

Regents' Testing Program
The University System of Georgia
Box 868
Georgia State University
University Plaza
Atlanta, GA 30303
(404) 658-4240

Major Fields

Beyond the major tests of general education and basic skills described above, there are nationally developed tests designed to assess knowledge and skills in major fields of study. Before describing these various instruments, let me again urge you to systematically compare tests with the objectives of your major programs. A given, commercially available test may not reflect what a particular department is trying to do.

If a department is primarily interested in assessment for program evaluation, it may not need to administer outside tests. Rather, it may be able to use the test results its students and graduates ordinarily provide in their application for graduate or professional education, or for licensure or certification. A post-graduation examination frequently taken by graduates from a given department will have obvious leverage with the department's faculty. Departments often develop "batting averages" out of such information.

State colleges and universities in Tennessee operate under a "performance funding" formula, with significant attention to the performance of students in majors for purposes of evaluating overall institutional effectiveness. This has forced the University of Tennessee System, the State Board of Regents, and the Tennessee Higher Education Commission to agree on examinations that institutions can use to assess the performance of major programs. The Tennessee Higher Education Commission has a list of approved tests for both baccalaureate and associate degree programs. The approved tests have been reviewed by relevant faculty and governing and coordinating board staffs. For a list of these test and more information on how they were developed and used, contact:

Tennessee Higher Education Commission
501 Union Building
Suite 300
Nashville, TN 37219-5380
(615) 741-3605

The Test Collection of ETS (1984) offers an extensive and detailed list of college-level achievement tests. This list includes equivalency tests, entrance examinations, certification tests, and achievement tests. The information provided for each test includes: an abstract description of the test and its purpose; the components within the overall test that assess particular skills or content; the ages and levels for which the test is suitable; and the organization that sells or distributes the test. To obtain a copy of ''Achievement Tests—College Level, December, 1984,'' write:

Test Collection
Educational Testing Service
Princeton, NJ 08541

The GRE Subject Tests are often used to assess, directly or indirectly, the knowledge and skills students have in their majors. According to the GRE 1984-85 *Subject Tests Descriptive Booklet* the primary purpose of subject area tests is:

> . . . . to help the graduate school admission committees and fellowship sponsors assess the qualifications of applicants in their subject fields. The tests also provide students with a means of assessing their own qualifications.
>
> Scores on the tests are intended to indicate students' mastery of the subject matter emphasized in many undergraduate programs as preparation for graduate study. (p. 3)

Tests designed to predict future performance in order to aid in the selection of candidates applying for admission to graduate or professional schools emphasize individual differences. As pointed out earlier, an emphasis on individual differences presents difficulties when the test is used for program evaluation. Nevertheless, student scores on such tests are frequently used whether formally or informally, to evaluate majors. Again, your respective departmental faculties will have to determine how the items of individual tests reflect major programs. Given the usual fee of $29 per test, it would be relatively expensive to have a significant number of students take this test for program assessment purposes.

Subject Tests are offered in biology, chemistry, computer science, economics, education, engineering, French, geology, history, literature in English, mathematics, music, physics, political science, psychology, sociology, and Spanish. To consider these tests, you should get a copy of *GRE Subject Tests Descriptive Booklet* as well as *GRE: Guide to the Use of the Graduate Record Examinations Program, 1984-85*, from:

Graduate Record Examinations
CN 6000
Princeton, NJ 08541-6000
or call:
Princeton, NJ (609) 771-7670
Berkeley, CA (415) 849-0950

Another set of examinations in which you may be interested is the ACT Proficiency Examination Program (PEP). These examinations were originally designed for the External Degree Program of the Board of Regents of the University of the State of New York. Outside of New York State, they are administered by ACT. These examinations are designed to assess proficiency in specific academic areas for the award of college credit; they range in testing time from three to seven hours, and in cost from $40 to $235 each. There are examinations in the following areas: Arts and Sciences (11 subjects); Business (18 subjects); Education (4 subjects); Nursing—associate level (8 subjects); Nursing—baccalaureate level (8 subjects).

The PEP Examinations are designed to reflect the content of individual courses rather than programs. Therefore, they will be of limited value in overall assessment of major programs, and it would be both administratively awkward and expensive to use these examinations for program assessment. For further information, contact:

Proficiency Examination Program
ACT
2201 North Dodge Street
Box 168
Iowa City, IA 52243
(319) 337-1000

Earlier I mentioned the CLEP Examinations, but there are also 33 Subject Examinations in CLEP. The basic purpose of the Subject Examinations is to assess proficiency in lower-division college courses for the purpose of awarding credit. Each examination requires 90 minutes. Some of the examinations have optional free-response or essay tests. The usual fee for each test is $30. Again, without exceptional circumstances, the CLEP Subject Examinations will be administratively and financially difficult to administer to groups of students for program evaluation purposes. They are not designed to reflect the comprehensive proficiency expected of a graduating senior in a major field.

Although originally developed for military personnel, ETS offers the DANTES (Defense Activity for Non-Traditional Education Support) achievement tests to colleges and universities for use with civilian students seeking college credit by examination. DANTES and CLEP cover different subject areas; for example, DANTES offers technological tests. Generally, the DANTES tests cover only the equivalent of one semester's work. Institutions can order DANTES tests and administer them at their convenience; the cost is $25 per test. The tests are untimed and take about 90 minutes each to administer. ETS scores the answer sheets. DANTES tests cover the following areas: Science (9 subject tests); Social Science (11 subject tests); Business (7 subject tests); Applied Technology (14 subject tests); Languages (4 subject tests); Mathematics (7 subject tests). If you are interested in reviewing the DANTES program, contact:

DANTES Program Office
P-166
Educational Testing Service
Princeton, NJ 08541
(609) 734-5212

IV. Local Assessments

After this discussion of externally available tests, we need to consider the development of assessment procedures and tests on your campus. While there have been serious efforts to improve instruction and to develop faculty as more effective teachers, little has been done to improve evaluation and testing. From his British experience, Heywood (1974) observed that:

> Examinations are the great afterthought of the educational process. Most new courses are set up without one thought being given to the methods of examining. (p. 2)

I believe improvements in instruction begin with feedback on student achievement. Such feedback is dependent on assessment, and the occasional use of outside, commercial tests is not enough. The best hope lies in encouraging faculty to improve their assessment procedures and to relate assessed student performance to program and instructional improvements.

Course Examinations

One place to begin a renewed concern for assessment is in course examinations. One might begin by asking that all faculty proposals for new courses include a final examination or some other summative assessment. Most faculty can write final examinations more easily than they can write specific course objectives. But by asking that tests or other means of assessment be included in proposals for new courses, faculty are more likely to define the outcomes of courses than if they are only asked to state objectives. I would also suggest another step: send new course proposals to two or three faculty members at other universities. Ask them to comment on the content and level of performance required of students in the proposed test or the alternative means of assessment, e.g. project, recital, etc. If this is done, the level of performance will have to be specified.

As McKeachie (1978) and Milton and Edgerly (1976) have helpfully demonstrated, one of the surest routes to improving collegiate instruction is by improving testing within courses. Good tests reflect course goals and content and give students feedback on their achievement. Warren (1984) has persuasively described processes for the collaborative development of tests and has experimented for many years with these processes in different kinds of postsecondary institutions. If you want to request his papers or seek his advice, contact him as follows:

Jonathan Warren
Research in Higher Learning
2360 Eunice Street
Berkeley, CA 94708
(415) 528-8414

Program Tests

You may decide in some cases to develop your own test to assess certain areas of general education or major fields. In many cases, you and your faculty will not be able to find externally available tests that reflect the particular emphases of your curriculum.

As you consider this possibility, you might consider Trudy Banta's approach at the University of Tennessee, Knoxville. Banta is helping faculty in several departments develop tests to assess major field proficiency when suitable "national" tests cannot be found. You may be interested in her "Plan for Comprehensive Test Development," to manage the on-campus construction of examinations to assess major programs. You may contact Banta at:

Learning Research Center
University of Tennessee, Knoxville
1819 Andy Holt Avenue
Knoxville, TN 37996-4350
(615) 974-2459

If you anticipate developing several tests on your campus, consider taking the following steps:

1. Develop a common procedure by which they are developed, reviewed, and approved.
2. Identify a test design consultant from your faculty who can develop the above procedure and who can work with faculty groups as they write and field test the examinations.

3. Have the test reviewed for content by at least two off-campus faculty acknowledged as experts by your faculty, and for psychometric quality by someone competent in the development of tests or other assessment procedures.
4. Provide test security.

Other Examiners

In "The Crisis of Integrity," cited earlier, O'Neill goes to the heart of the problem: the same person who teaches the student, also tests and certifies the student. In this, American higher education is different from European and British education. There are ways, however, that we could use other examiners in addition to the student's primary instructor. If, for example, a senior comprehensive in each major is required, more than one faculty member could be involved in evaluating student papers, projects, or examinations. Alumni with some graduate work or demonstrated professional expertise related to a particular major could be used on a team to evaluate performances or products in senior comprehensives.

Improvement of instruction is tied to re-establishing a sense of pride-in-craftsmanship in instructors. Craftsmen identify with their products, and craftsmanship is reinforced by the response of purchasers and informed observers. When a faculty member in a given department presents the work of a senior major to two or more colleagues from within or outside the campus, there is an opportunity to receive the type of evaluation that engenders pride of craftsmanship.

Senior Comprehensives

Of all the initiatives one might take to encourage assessment of outcomes, I would begin with senior comprehensives. Sometime in the senior year, each major should complete a major paper or project under the guidance of a faculty member in that department. That paper or project would be judged in some predetermined, systematic way by two or more persons deemed by the department faculty as competent to appraise summative undergraduate work in the field.

Such comprehensive papers and projects should require a student to demonstrate not only knowledge and skill of his major, but much of his general education. Senior comprehensives are not as common as they once were, but a number of institutions still have them at least in some departments. I hope they will again become rather common, and that regional accreditation agencies will require members of visiting committees to review student work produced in them.

Swarthmore College has had an external examination system as part of its honors program since it was established by President Frank Aydelotte in 1922. A student reads for Honors in his/her junior and senior year, preparing to take four examinations in his/her major and two in his/her minor. External examiners (faculty from other institutions) evaluate students' three-hour written examinations, and, in addition, come onto the campus to conduct an examination of each student.[4]

V. Assessing Attitudes and Behaviors

While the primary focus in outcomes assessment is on academic achievement, we remain interested in the attitudes and behaviors affected by the campus experience. This section presents a very brief overview of some ways to observe or assess student attitudes and behaviors.

Questionnaires can provide self-reported information about student values, interests, beliefs, and behaviors. If your institution includes in its mission having an effect on student attitudes and behaviors, you will need ways to collect reliable and valid data about them. Observations and inventories are two basic ways to get such data.

Observations

One can learn a great deal simply by observing behavior. You might ask the anthropologists and sociologists on your campus to help you identify relatively unobtrusive and inexpensive ways to observe and record campus behaviors related to the campus' particular values. We can infer much about values and changes in values from students' entertainment choices, community service, campus religious life, dress patterns, fraternity and sorority activities, involvement in political issues and activities, and numerous other social behaviors. Campus social scientists, teamed with campus journalists, could effectively keep your campus community informed about behavior patterns and their inferred meanings, without an overemphasis on formal surveys.

Inventories and Questionnaires

Cronbach (1960) referred to questionnaires and inventories of attitudes and behaviors as "tests of typical performance." The purpose of typical performance assessments is to determine what one *usually* feels, believes, or does. They contrast with tests of ability and achievement designed to reflect *maximum* performance (see Cronbach, pp. 29-31). In maximum performance tests, one is supposed to do his or her best.

A maximum performance test of composition would require writing an essay to be judged for punctuation, grammar, and organization. A typical performance assessment would be reviewing the punctuation, grammar, and organization of a sample of letters randomly selected from the routine correspondence of an office.

In assessing beliefs, values, and attitudes, we want to know how one actually feels, as opposed to how one believes he or she *should* feel. Responses to typical opinion polls and questionnaires are vulnerable to influence from one direction or another. Therefore, the questions must be worded to minimize bias. Usually people respond more candidly if responses are anonymous. Finally, knowledge and skill achievement tests may focus on individual as well as group performance. In contrast, reports of responses to inventories and behaviors should focus only on groups.

Pace

One self-report inventory is the College Student Experiences questionnaire developed by C. Robert Pace. For the most part, this questionnaire asks students about college-related activities in which they have actually engaged, i.e., use of libraries, interaction with faculty beyond the classroom, involvement in the arts, etc. You may want to review this inventory, along with Pace's in-depth discussion, in *Measuring The Quality of College Student Experiences*:

C. Robert Pace
Higher Education Research Institute
Graduate School of Education
UCLA
Los Angeles, CA 90024

As you consider using questionnaires and observations to determine what effects your institution is having on students, I suggest you also review Pace's *Measuring Outcomes of College* (1979). Pace has designed and conducted many surveys of alumni, so you will find the chapter "Achievement After College: Alumni" (pp. 48-113), very helpful.

ACT

ACT currently offers eleven surveys to assist institutions in evaluation: Adult Learner Needs Survey; Alumni Survey; Alumni Survey (2-year College Form); Entering Student Survey; Student Opinion Survey; Student Opinion Survey (2-year College Form); Survey of Academic Advising; Survey of Current Activities and Plans; Survey of Postsecondary Plans; Withdrawing/Nonreturning Student Survey; Withdrawing/Nonreturning Student Survey (Short Form).

I have recently used The Alumni Survey and the Student Opinion Survey in an accreditation self-study. The Alumni Survey elicits information about the respondents' background, continuing education, college experiences, and employment history, along with space for thirty additional local questions. The Student Opinion Survey covers the respondent's background, evaluation of college services, and satisfaction with college environment, as well as providing thirty spaces for additional questions, and write-in spaces for comments and suggestions. They are easy to administer, and the scored responses are reported in an easily understood format. If you are interested in these surveys, contact:

Institutional Services Area
ACT
2201 North Dodge Street
Post Office Box 168
Iowa City, IA 52243
(319) 337-1102

ETS

ETS also offers eight surveys for institutional evaluation: Institutional Goals Inventory; Community College Goals Inventory; Small College Goals Inventory (there is also a Canadian Institutional Goals Inventory and a Spanish/English Institutional Goals Inventory); Student Instructional Report; Institutional Functioning Inventory; Student Reaction to College; Program Self-Assessment Service; Graduate Program Self-Assessment Service.

In an institutional self-study, I have used the Small College Goals Inventory (SCGI) and the Graduate Program Self-Assessment Service (GPSAS) questionnaires along with the Undergraduate Program Self-Assessment Service (PSAS) questionnaires. The SCGI allows a variety of constituents, students, faculty, alumni, board members, etc., to compare what *are* and what *should be* the institution's goals. The PSAS provides different questionnaires for enrolled students, alumni, and faculty to evaluate departmental programs and elicits responses in sixteen areas, including environment for learning, student accomplishment, and student satisfaction with the program. Space is provided for twenty additional local items. As with the ACT instruments, these surveys are easy to administer and responses are reported so as to be easily interpreted by faculty. For more information or for examination copies, contact:

College and University Programs
Educational Testing Service
Princeton, NJ 08541
(609) 734-1162

Values Inventories

Some institutions are particularly interested in detecting shifts in the values of their students during their campus experience. For brief overviews of research on the effects of college on student values, see Pace (1979), Astin (1977), Bowen (1977), Winter, McClelland, and Stewart (1982), and Feldman and Newcomb (1969). You can make interesting comparisons with value inventories:

1. Freshman-to-senior changes in values.
2. Students, faculty, administration, and staff similarities and differences in values.
3. Changes in the values of new freshman classes from year to year.

For the last of these, you might consider participating in the Cooperative Institutional Research Project's (CIRP) Annual Survey of American College Freshmen, which, since 1966, has included a significant section on values. More than 600 institutions currently participate in CIRP. If you are not one of them, write or call:

Cooperative Institutional Research Program
Graduate School of Education
UCLA
405 Hilgard Avenue
Los Angeles, CA 90024
(213) 825-1925

If, on the other hand, you wish to develop a local values inventory, I suggest that you first review:

Study of Values
Gordon W. Allport, Philip E. Vernon, and Gardiner Lindzey
The Riverside Publishing Company
Post Office Box 1970
Iowa City, IA 52244
(319) 354-5104

Rokeach Value Survey
Milton Rokeach
Halgren Tests
873 Persimmon Avenue
Sunnyvale, CA 94087
(408) 738-1342

Institutional Use

What practical use can be made of questionnaire data? They provide a beginning point from which relevant groups of faculty, administrators, and students can discuss the effects of programs. That is, do not take the tabulation of survey results as "reality." They are no more reality for the institution than a vocational interest inventory is the reality of a given student's career goals. The individual student's responses to an interest inventory provide him or her and the counselor a basis for their discussions. Surveys are best used in organizational development, as pump-primers for discussion and further investigation. Obviously, longitudinal studies of changes in attitudes and behaviors are preferable for these purposes, and must be planned to stretch over at least four, and probably six, years.

VI. Assessment Centers

Up to now, I have dealt with more and better uses of tests, inventories, and other assessment procedures with which most of us are aware. While you are probably not immediately interested in radically different arrangements for assessment, I believe one should anticipate nurturing a climate that will eventually support assessment as more than an add-on to the current intra-course, teaching/testing system. If approached with a combination of the following mutually supporting commitments and services, assessment can facilitate educational renewal:

1. Granting credit on the basis of demonstrated achievement;
2. Identifying and using competent third-party examiners;
3. Stating clear, expected achievements in general education and major programs;
4. Integrating as much as possible the roles of "instructor" and "academic adviser" into the one role of "mentor;"
5. Developing a comprehensive and integrated student advising, testing, educational, and career counseling service.

Assessment centers originated not in colleges but in corporations and the military. Thornton and Byham (1982) describe them as follows:

> An assessment center is a comprehensive, standardized procedure in which multiple assessment techniques such as situational exercises and job simulations (i.e., business games, discussion groups, reports, and presentations) are used to evaluate individual employees for various purposes. A number of trained management evaluators, who are not in a direct supervisory capacity over the participants, conduct the assessment and make recommendations regarding the management potential and developmental needs of the participants. The results of the assessment are communicated to higher management and can be used for

> personnel decisions involving such things as promotions, transfer, and career planning. When the results are communicated to the participants, they form the basis for self-insight and development planning. (p. 1)

Moses (1977) describes the assessment techniques commonly used as—

> . . . group exercises, business games, in-basket exercises, pencil-and-paper tests, and interviews. They may also include specially designed role-playing problems, phone calls, or simulated interviews. (p. 4)

Moses also identified three general characteristics of successful assessment centers:

1. Assessors were quite familiar with the job or duties they were assessing.
2. Simulation exercises are used more than pencil-and-paper tests.
3. They made predictions about specific outcomes rather than personality traits or individual characteristics.

In contrast, the less successful centers "relied heavily on tests rather than simulations and made descriptions of personality traits rather than predictions of specific behaviors." (p. 9)

Alverno Adoption

Alverno College provides a well-known instance of a successful assessment research, development, and service center. Ewell (1984) describes Alverno's very different, and somewhat complex, assessment-based program briefly and clearly. The Alverno approach is described in more detail in *Assessment at Alverno College* (1979), written by the Alverno faculty.

The Alverno curriculum is designed to help each student demonstrate the following eight general abilities: effective communications ability; analytical capability; problem solving ability; valuing in a decision-making context; effective social interaction; effectiveness in individual/environmental relationships; responsible involvement in the contemporary world; aesthetic responsiveness.

Each student must demonstrate competence at six levels in each of these abilities. The types of required abilities and levels of performance are not classroom-bound, nor are they all amenable to conventional paper-and-pencil tests. To produce relevant assessment procedures and an organizational unit to develop, refine, and administer the assessment procedures, Alverno had to look for help beyond the academy and the national testing agencies. They found a paradigm to adopt and adapt in the AT&T assessment center program.

The core purpose of assessment at Alverno is feedback for the development of individual students. *Assessment at Alverno College* notes, ". . . the ultimate *raison d'etre* for assessment is to provide the student, at each of many steps in her development, with progressively fuller and more individual profiles of her emerging combination of gifts, skills, and styles, so that she can become an independent learner" (p. 7). Mentkowski and Loacker further describe that function as follows:

> Whether it is as simple as a series of one-paragraph responses to questions about a film, or as complex as presenting a park-use plan to a neighborhood association, faculty try to use each assessment situation as a learning experience. Ideally, assessment should contribute to and culminate a process of working toward explicit, known goals, with frequent stops to find out "the state of the art" in the ability that the student is working to develop.[5]

The paper by Loacker, Cromwell, and O'Brien in this volume presents a fine elaboration of this process, but if, in addition, you want the Alverno documents to which I have referred, write or call:

Alverno Productions
Alverno College
3401 South 39th Street
Milwaukee, WI 53215-0001
(414) 647-3780

Relatively small institutions whose mission, like that of Alverno, focuses principally on student learning and development rather than research and publication, and which are bound together by a strong ethos, have a great opportunity in assessment. But others should be wary of taking an Alverno assessment "blue print" and setting it up *in toto* if their campus culture is not characterized by functionally common goals, a familial-like organization, and sub-units within the institution with which individuals identify.

Other Approaches

Some organizations have established instructional improvement centers to assist faculty in the systematic design and media support of course instruction. In similar fashion, an island of test development expertise could be established at any college or university. It could assist faculty in practical matters such as building computerized test item banks and using media in testing. As the assessment center establishes credibility through practical service, it will become involved in assisting faculty in the basic design of tests particularly for large-enrollment, multi-section courses.

Your teacher education program is probably one of the places where an assessment center may prove very effective. One of the reasons corporations establish assessment centers in business and industry is to select individuals for further training, development, and promotion. Fairness and profitability demand that the assessments be job-related. So the assessments must be lifelike. Similarly, the abilities of future teachers to cope with a variety of real-life teaching circumstances and dilemmas should be assessed in ways as closely approximating real classrooms as possible. I believe many of the same techniques used in business and industrial assessment centers could be adapted for teacher education. The technology and hardware usually associated with instruction can be used effectively in assessment. Obviously, there is no reason to restrict assessment centers to teacher education; I just happen to believe it is particularly needed there. It has great potential in many areas as diverse as nursing, business administration, music, art, engineering, or any major field of study.

Most lasting changes are "grown out" slowly rather than imposed. Given this perspective, you may want to get a few faculty members together to explore and discuss the adaptation of assessment centers to colleges. Help them examine the assessment center literature; arrange for some of them to visit assessment centers in corporations and at other colleges and universities with assessment centers. From these explorations, you could "grow out" an Assessment Center which would reflect your particular curriculum and circumstances.

Conclusion

Hopefully, I have hit on some features of assessment you can use immediately and some that will mean more after you have been at it awhile. I suggest that you not wait to start institution-wide assessments until the "perfect" test or inventory is found or developed. Starting with something, realizing its imperfections, and being appropriately tentative with its results is a far more productive strategy than talking the issue to death. After all, our principal interests in assessment are twofold:

1. Making sure students' achievements are commensurate with the credits and degrees we award them.
2. Getting information that will stimulate and guide the improvement of instruction and curricula as well as the personal development of students.

The integrity of the credentials we award are at stake in the first interest, and the integrity of our academic life in the second.

The corporate and government worlds that provide the capital on which we exist, and which hire most of our graduates, are thinking quality assurance and will expect us to do the same. We should not use the walls of academic freedom to shield low standards and ineffective instruction. We should use the current national interest in quality as an opportunity to assure standards and improve instruction. This can be done consistent with the best academic tradition and practice, particularly if we include the general procedures historically associated with "external or third party examiners."

As you move into assessment, I suggest you find someone to observe and comment on the "organizational development" implications of what you want to do. An emphasis on assessment will affect the way your institution functions as an organization. You not only need help in the technical side of assessment but also in the nurturing of a climate characterized by deep concern for ''results'' over ''form,'' commitment to high standards, and concomitant interest in helping students reach those standards. In the last analysis, an emphasis on assessment is more of an *attitude* than a collection of tests. Attitudes, as you know, cannot be mandated from the top down, but are nurtured from the bottom up.

Notes

1. Joseph Hammock. Personal communication with author.
2. For more complete discussions of the difficulty of using selection-referenced tests to assess effects of different instructional treatments, see: Joseph Hammock, "Criterion Measures: Instruction vs. Selection Research." Presented at the annual meeting of the American Psychological Association, September, 1980; Robert Glaser, "Instructional Technology and the Measurement of Learning Outcomes: Some Questions," *American Psychologist*, 1963, 18, 519–521; and Robert Glaser and David J. Klaus, "Proficiency Measurement: Assessing Human Performance," in *Psychological Principles in System Development*, Robert Gagne, editor. New York: Holt, Rinehart, and Winston, 1962, 419–474.
3. See Aubrey Forrest and Joe M. Steele, "Defining and Measuring General Education Knowledge and Skills—COMP": *Technical Report* 1976-81, The American College Testing Program, 1982.
4. Swarthmore College Bulletin, 1985-86, 44–48.
5. From the manuscript, "Assessing and Validating the Outcomes of College" which is to be the fourth chapter of *New Directions for Institutional Research: Assessing Educational Outcomes*, edited by Peter Ewell.

References

Academic Preparation for College: What Students Need to Know and Be Able to Do. New York: The College Board, 1983.

Assessment at Alverno College by the Alverno College Faculty. Milwaukee: Alverno College, 1979.

Astin, A.W. *Four Critical Years*. San Francisco: Jossey-Bass, 1977.

Bowen, H. *Investment in Learning*. San Francisco: Jossey-Bass, 1977.

Branscomb, H.; Milton, O.; Richardson, J.; & Spivey, H. *The Competent College Student: An Essay on the Objectives and Quality of Higher Education*. Nashville: Tennessee Higher Education Commission, 1977.

Dressel, P.L. *College and University Curriculum*. Berkeley: McCutchan Publishing Corporation, 1968.

Ewell, P. *The Self-Regarding Institution: Information for Excellence*. Boulder: National Center for Higher Education Management Systems, 1984.

Feldman, K. and Newcomb, T. *The Impact of College on Students*. San Francisco: Jossey-Bass, 1969.

Forrest, A., & Steele, J.M. *COMP: Defining and Measuring General Education Knowledge and Skills* (Technical Report 1976-81). Iowa City: The American College Testing Program, 1982.

Hammock, J. "Criterion Measures: Instruction vs. Selection Research." Paper presented to the American Psychological Association, September, 1960.

Harris, J. "Baccalaureate Requirements: Attainments or Exposures?" *The Educational Record* 53 (Winter, 1972), 59–65.

Heywood, J. *Assessment in Higher Education*, New York: John Wiley & Sons, 1977.

Integrity in the College Curriculum: A Report to the Academic Community (Project on Redefining the Meaning and Purpose of Baccalaureate Degrees). Washington: Association of American Colleges, 1985.

McKeachie, W.J. *Teaching Tips: A Guidebook for the Beginning College Teacher*. Lexington, Massachusetts: D.C. Heath and Company, 1978.

Mentkowski, M., & Loacker, G. "Assessing and Validating the Outcomes of College," in P.T. Ewell (ed.) *Assessing Educational Outcomes*. New Directions in Institutional Research No. 47. San Francisco: Jossey-Bass, pp. 48–64.

Milton, O. & Edgerly, J.W. *The Testing and Grading of Students*. Change Magazine and Educational Change, 1976.

Mingle, J.R. "Measuring the Educational Achievements of Undergraduates: State and National Developments." Denver: State Higher Education Officers, January, 1985.

Morris, L.L., & Fitz-Gibbon, C.T. *How to Measure Achievement*. Beverly Hills: Sage Publications, 1978.

Moses, J.L. "The Assessment Center Method." *Applying the Assessment Center Method*, edited by J.L. Moses and W.C. Byham. New York: Pergamon Press, 1977.

O'Neill, J.P. "The Crisis of Integrity." Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada, April, 1983.

Pace, C.R. *Measuring Outcomes of College*. San Francisco: Jossey-Bass, 1979.

Scannell, D.B. & Tracy, D.B. *Testing and Measurement in the Classroom*. Boston: Houghton Mifflin Company, 1975.

Thornton III, G.C., & Byham, W.C. *Assessment Centers and Managerial Performance*. New York: Academic Press, 1982.

Troutt, W.E. "Unbundling Instruction: Opportunity for Community Colleges," *Peabody Journal of Education*, 56, 4 (July, 1979), 253–259.

Wang, W.K.S. "The Unbundling of Higher Education." *Duke Law Journal*, 53 (1975), 53–90.

Warren, J. *Meeting the New Demand for Standards*. No. 43 in *New Directions in Higher Education*. San Francisco: Jossey-Bass, December, 1984.

Winter, D.G.; McClelland, D.C.; & Stewart, A.J. *A New Case for the Liberal Arts*. San Francisco: Jossey-Bass, 1982.

The Costs of Assessment

by Peter T. Ewell
and Dennis P. Jones

The report of the Study Group on the Conditions of Excellence in American Higher Education, *Involvement in Learning: Realizing the Potential of American Higher Education* (1984), identified assessment and feedback as one of three conditions for achieving excellence in undergraduate education. The Study Group argued that "institutions should be accountable not only for stating their expectations and standards but for assessing the degree to which those ends have been met." (p. 21) The underlying theme is that acquiring and using information about performance is a necessary ingredient in any attempt to foster learning and self-improvement. We all recognize the legitimacy of this argument when applied to students. Most accept the notion that such evaluations need to be formative as well as summative. The assessment process may be implemented badly at times, but there is substantial agreement that the evaluation of student learning and development ought to guide the teaching and learning process.

Both the Study Group and Ewell (1984) go further. They argue that what holds true for assessing students also holds true in a broader context. Specifically, they maintain that the road to improvement of courses, programs, and indeed the institution itself, involves regularly collecting information on institutional and program effectiveness, and using such information as the basis for improvement. Intellectually and conceptually, the argument has the ring of reason. We can readily accept the notion that information is knowledge, and that we ought collectively to be more knowledgeable about our institutions and the programs they house. Ergo, assessment information about institutions and programs, as well as individual students, is desirable.

On a more practical level, however, the recommendations of *Involvement in Learning* with regard to assessment and feedback are often greeted with skepticism. Indeed, the level of skepticism itself is revealing; it stems primarily from unfamiliarity rather than from unfortunate experience. The skepticism that we have observed surfaces in the form of two concrete questions. First: "Can assessment actually be accomplished; is it feasible?" As a technical question, this is being answered in the affirmative, supported by a growing body of institutional experience with wide-ranging assessment programs. But the second question is equally pragmatic: "How much does it cost?" The underlying tone of the question reflects a conviction that the costs are high.

We address the latter question in this paper. In the following section, we present a simple conceptual schema to delimit the dimensions of the question. In the balance of the paper, we present estimates of the costs of assessment for different types of institutions.

An Analytic Framework

To properly address the costs of assessment, we must pose and answer two distinct questions. The first question is "the costs of assessing what?"—a question of unit of analysis. The second question is "what costs?"—a question of what to count. These two questions are treated separately below.

Unit of Analysis

The unit of analysis with which we are most traditionally comfortable in assessment is the individual student. In the normal course of events, the individual student experiences a wide variety of assessments in the process of being admitted to, and making progress through, an institution. Students take ACT and SAT tests as part of the application process. Incoming freshmen commonly take a battery of institutional tests for placement purposes immediately upon arriving at campus. Most pervasive of all assessment activities are the many tests that students take in each and every course in which they are enrolled. By such means, we collect mounds of assessment data on students. Our facility for turning this data into information, however, remains limited. But we do at least gain enough information from these activities to convince ourselves that individuals do or do not deserve to be certified as academically worthy and eligible to receive a degree, diploma, or certificate.

Beyond the individual student, the units of analysis with which we are primarily concerned are the program or curriculum, and the institution as a whole. With regard to individual programs or curricula, assessment questions abound. A central question is: "Are the students who have completed the program emerging with the intended level of knowledge and skills, and are they proceeding to fill intended roles in desirable ways?" Corollary questions include the attractiveness of the program to particular groups of students and student satisfaction with the educational experience provided by the program. Each of these questions can be illuminated by periodic assessment of the outcomes of the program. While many of the basic data needed to address these questions are the same as those needed to assess individual student development, the ways in which we analyze these data will be different. For program evaluation, the primary need is to look at the *collective* performance of a particular body of students (or a representative sample thereof). This means not only examining mean or median performance, but also investigating and accounting for the nature of variations around these central tendencies.

Finally, comprehensive assessment requires information about the performance of the institution as a whole. It is at this level that questions of feasibility become most widespread and acute. As a consequence, it is at this level that assessment is least frequently conducted. Given a wide array of outcomes attributable to almost any college or university—and given that the typical institution tends to claim credit for contributions to growth along all of these dimensions—there is an understandable inclination, in Kenneth Mortimer's words, to "measure everything that moves." With this perspective, it is easy to see how questions of cost emerge as a real issue.

To estimate the costs of assessment, we first deal with the appropriate scope of assessment. With this requirement in mind, we want to emphasize that the essence of institutional assessment is to "measure your mission." Adherence to this simple principle can help insure that institutional assessment is a carefully focused activity. Posing the question in this way also requires that assessment be tailored carefully to reflect the distinctive aspects of each institution. If the institution in question is primarily oriented toward professional and occupational training, appropriate assessment should be focused upon the documented success of graduates in the professions and occupations for which they were trained. For liberal arts colleges, in contrast, primary assessment strategies tend to examine student development along dimensions of general knowledge and general skills. In major research universities, assessment may be concentrated on student success in the major. There are, of course, variations on all of these themes, including consideration of student satisfaction with the experience, as well as educational "value-added."

The Costs Considered

There are innumerable concepts of, and ways to, calculate costs. Among them are direct costs, indirect costs, full costs, average costs, marginal costs, and opportunity costs. The appropriateness of each of these approaches to costing is determined by the use of the resulting information. Consequently, the real question for us is not simply "What is the cost of assessment?" Because the issue is usually raised in a managerial or resource allocation context, our question becomes, "How much *more* money do we have to spend to put in place an assessment program that is appropriate to our needs?" Using this notion as a guide, we have passed over attempts to estimate the cost of student assessments already undertaken as a regular part of the student's coursework. While it might be possible to calculate the actual proportion of faculty instructional effort already attributable to in and out of class assessment activities, this would yield information without a purpose. It is more important to attempt to determine the level of *regular* investment the institution must make in addition to these ongoing activities. This is an incremental or marginal cost. Certainly we recognize that the dollars currently being spent for assessment can often be spent more effectively, and that assessment programs can often be improved at no added cost. Such reallocation issues, however, are not within the domain of this paper.

Putting these two dimensions together results in a matrix that displays major cost considerations as follows:

Dimensions of the Topic

Unit of Analysis \ Costing Focus	(Additional) Incremental Costs	Current Costs
Institution	▨	
Program	▨	
Student		

The portion of assessment costs we discuss in the balance of this paper is indicated by cross hatched areas in this diagram. In the following sections, we provide estimates of typical incremental costs for establishing and maintaining institutional and program level assessment programs. At best, these estimates are exceedingly rough. In spite of their limitations, however, they do provide reasonable ballpark figures regarding the level of costs that might be expected by an institution embarking on a comprehensive assessment program.

Estimating Costs of an Institutional Assessment Program

Estimating the actual new costs of establishing an assessment program at a given college or university can be a complex undertaking. As most institutions already collect some data on student achievement and program effectiveness, creating a comprehensive assessment program may thus involve coordinating a number of activities for which the institution has already paid. An additional difficulty is the level of analysis at which assessment takes place. While data gathering on institutional effectiveness remains relatively rare, all institutions collect some data on individual student performance. As noted above, the kinds of data routinely collected on individual students at most campuses may or may not be consistent with good pedagogy. But in general, changing assessment methods and policies at this level will not entail significant additional costs.

Because of these difficulties, several caveats are necessary before we present some actual cost estimates. First, we will base our estimates primarily on direct costs—those costs incurred by fielding new test and survey instruments, and by making use of the results. While a variety of indirect or overhead costs might be considered (for example, professional time spent drawing the implications of assessment results, faculty and administrative time spent reviewing programs in the light of assessment data, and the like), these will vary so greatly that concrete estimates would be problematic.

Our second assumption is that an institution will adopt an explicit *program* for assessing instructional effectiveness. This means that various related instructional evaluation efforts will be centrally coordinated and supported by a staffed, visible office. Establishing such an assessment program, it is important to note, may involve considerable reallocation of existing, funded functions. For example, many institutions already fund a testing center, an institutional research office, or an academic planning office. Functions of each of these existing offices are commonly included in a comprehensive assessment program. Furthermore, many individual data gathering efforts included in assessment programs may already be in place in one or more of these locations. Many institutions, for example, regularly administer student surveys, such as the Cooperative Institutional Research Program (CIRP), or conduct surveys of students upon graduation or withdrawal. Many institutions regularly assess student abilities for placement purposes on entrance. Finally, many institutions regularly administer professional or pre-professional certification tests that assess particular skills gained in the course of instruction.

Our final assumption is that assessment will rest primarily on traditional "paper-and-pencil" testing and survey methods. Certainly there are many alternatives to cognitive tests and forced-choice survey instruments, and they should be carefully considered in building an assessment program. Use of external assessors drawn from the local business and professional community, as practiced by Alverno College and others, constitutes one such alternative (Mentkowski and Doherty, 1983). Traditional jury or panel ratings of performances in such fine arts disciplines as music, drama and dance provide another. In such cases, assessment costs can be estimated in terms of the professional time committed by external evaluators. The issue of costs is far more difficult when faculty themselves play these roles—either in addition to, or instead of, traditional grading practices. In such cases, the "costs of assessment" can easily be viewed as part of an individual faculty member's existing assignment. Because of these difficulties, we consider only the direct costs of more traditional assessment methods in the discussion that follows.

Cost Elements for Assessment Programs

In constructing assessment programs, most institutions incur costs in four basic areas. First, assessment instruments (tests and surveys) must be constructed locally or purchased from an outside vendor. Second, these instruments must be administered to students. Third, the resulting data must be analyzed and disseminated. Finally, the assessment effort itself must be coordinated. Each of these costs is driven by different parameters, and by the kinds of choices that institutions may make within each cost element.

1. Instrument Costs—Various cognitive tests and student surveys form the basis for any assessment data gathering effort. Before they can be fielded, tests and instruments must first be developed or obtained. If they are developed locally, costs are incurred by faculty and measurement specialists in constructing the test or survey. After initial development, such instruments can be produced on a regular basis—generally at lower cost than comparable commercial instruments.

The alternative to constructing instruments locally is to make use of commercially available tests and surveys. Examples include the Graduate Record Examination (GRE) Field Examinations—often used as senior assessments of knowledge in the major field, various professional and pre-professional certification and placement tests (for instance, the National Teacher Examination) used for the same purpose, and course content examinations such as the College Level Examination Program (CLEP)—designed for awarding credit-by-examination, but increasingly used to assess mastery of lower-division basic skills. Examinations such as these are obtained through purchase—generally on a per-instrument basis.

Because of the difficulties involved, relatively few institutions choose to design their own cognitive tests. Generally, local achievement tests are developed as senior assessments in fields not currently covered by existing commercial instruments. More rarely, colleges have developed their own general education assessment instruments. Local examinations have also been developed because faculty feel that existing commercial instruments do not adequately cover the field as taught in their own curricula. Developing good subject area examinations can be a time-consuming exercise and additional resources are required for pilot testing the instrument and for subjecting individual test items to careful review by testing/measurement specialists.

Institutions that have constructed such field examinations have usually treated their development as a departmental activity. This practice tends to bury many test-making expenses in ongoing departmental administrative costs associated with curriculum development and review. If the full cost of such activity were calculated, it would undoubtedly be quite high. In practice, however, budgeted test development costs tend to be treated as a short-term overload assignment for particular departments—covering part, but not all, of the resources required. For example, one major research university is currently undertaking development of twenty such departmental examinations at a budgeted cost of $2,000 each. This university judges the cost as an appropriate increment only because departments are expected to reallocate additional existing resources to test-making that are already "budgeted" for curriculum review and improvement activities.

Development of local surveys—either of currently enrolled students or of former students (graduates and dropouts)—is much more common than development of local cognitive assessment instruments. In general, good survey instruments can be designed for less than the costs associated with cognitive tests. Some economies result from the fact that many common models are available. References such as McKenna (1983), Pace (1975), and California Community Colleges (1984) provide excellent and accessible lists of items commonly included on student surveys.

Commercial tests and surveys are generally purchased on a per unit basis. For cognitive tests and examinations, the unit price includes scoring as well as the price of the instrument. Individual prices vary considerably from a low of $7/exam for instruments such as the ACT Assessment Entrance Examination, through $29/exam for the GRE, to a high of $43/exam for such instruments as the National Teacher Examination administered by ETS. In some, but not all, cases multiple purchase discounts are available for institutions.

Commercial student surveys are generally available for individual purchase, with or without associated processing and analysis services. Prices for individual instruments range from a low of 15¢/survey to approximately $1/survey. When analysis services are used, total costs average $3–$5 for each completed questionnaire. In addition, institutions can purchase a tape of responses for $40–$150 and can obtain comparative reports consisting of responses from other institutions that have used the instrument.

2. **Administration Costs**—Once in hand, tests and surveys must be administered to students. In some institutions, existing testing centers established for placement or diagnostic testing may bear some of this burden. In most cases, however, the number of instruments to be administered simultaneously will require resources beyond those available to the typical institutional testing center. Cognitive test administration is generally a straightforward, in-class exercise, but even so, considerable administrative costs may be incurred. For cognitive tests, proctors may need to be employed for multiple test locations.

For some types of tests (for example, the ACT College Outcomes Measures Project) special video and audio equipment must be available and operated. If special testing sessions are scheduled, students must be notified where they should appear, and follow-up procedures put into place to insure that they do in fact appear. Finally, costs will be incurred in recording results, and if desired, in sending test results directly to students.

Some of the same procedures will be typical of in-class survey administration. Generally, however, in-class student surveys will not require supervision, and survey questionnaires will take less time to complete than examinations (an average of 10 to 20 minutes as compared to the typical three-hour length of most examinations). Moreover, many student surveys can be administered in existing settings, for example, at student registration or during orientation programs. Because of the ready availability of such mechanisms, entering student questionnaires are the kinds of survey instruments most easily and directly administered to students.

For program graduates or withdrawing students, or for currently enrolled students who may be difficult to reach in an available "captive" setting, survey administration by mail will be typical. Mailed survey costs vary with the number of respondents to be reached, the number of mailings undertaken to maximize response, and the estimated response rate. In order to obtain acceptable response rates, most institutions use more than one mailing, and often supplement results with telephone follow-ups of non-respondents. Most sources recommend the use of first-class postage on both mailout and return envelopes (Dillman, 1982). Costs for recording and tabulating responses should also be included in any analysis. Based upon such parameters, typical costs for conducting mailed surveys will average $1.50–$3 per completed instrument.

3. **Analysis Costs**—As noted above, commercial cognitive tests include analysis and processing expenses with the cost of the instrument. Scoring and analysis services are also available for most commercial surveys. These services include costs for data entry, computer analysis, and production of a simple frequency or cross-tabulation report. In many cases, however, available data will need to be further analyzed for policy purposes. In the case of test data, individual student performance results may be correlated with student characteristics, with course-taking patterns, or with other elements of the institutional experience. This task entails creating data sets which make use of a variety of data elements beyond simple test performance. The same is true of student survey data. Tapes of questionnaire responses—generally available from the providers of the instruments—can also be analyzed locally using an available statistical package. All such exercises entail both personnel and data processing costs.

In the case of locally developed tests and surveys, analysis designs will have to be created from scratch. Like instrument design, this is a one-time cost, but it can be considerable. A set of analysis routines must typically be written using a standard statistical package (for example, SPSS or SAS) or using a common programming language.

Similarly, response coding schemes must be devised and, if applicable, machine scoring procedures using mark-sense equipment established. In the initial stages, considerable care must be taken to develop error-checking procedures and methods for handing missing, incomplete or contradictory information. Once such procedures are put in place, however, ongoing costs for data analysis will be minimal, involving personnel and computer time.

4. Coordination Costs—Establishing a comprehensive program of institutional assessment may require investments beyond the direct costs associated with procuring, administering, and analyzing a variety of data gathering instruments. Such comprehensive programs are centrally administered and involve coordinating many kinds of data collection and analysis activities. Indeed, the most effective of such programs are located in distinct, specially created offices—for example, Alverno College's Office of Research and Evaluation, and the University of Tennessee at Knoxville's Learning Research Center.

Costs associated with establishing an office of this kind include those for new professional and support staff, office space to house these personnel, and ongoing operating expenses. In estimating such costs, it is important to attempt to isolate the new functions that such offices will fulfill from those associated with the existing, previously funded activities that such offices will now assume. For example, overseeing an annual alumni survey effort and writing data reports on this activity may already be part of an institutional research responsibility and may be built into the cost structure of a new assessment office. Similarly, existing diagnostic testing and measurement activities may be folded into such an office's established responsibilities. Generally, however, such functions as administering comprehensive examinations in general education, or working with faculty to develop local survey and test instruments, are not covered in the institution's current cost structure.

In many cases, existing personnel are reassigned to provide staffing for an assessment center. Faculty with appropriate research backgrounds in the social and behavioral sciences, or Testing/Institutional Research professionals, may be taken from their current assignments and given responsibility for coordinating institution-wide assessment activities, for designing instruments, or for analyzing and interpreting test or survey results. In such cases, estimating costs may be difficult because the relevant question is the cost of *replacing* the reallocated staff member in his or her original function. Often this can be done with part-time instructors or research assistants at a cost that is probably far less than that of full replacement. In other cases, the reassigned person may be currently underutilized, and may consequently not need full replacement. Alternatively, however, the reallocated position may be in a high-demand area, and a premium must be paid for its replacement.

As a result of the extreme variation in current practice, any estimates of coordination costs will be approximate. In each of the cases discussed below, we attempt to disaggregate these costs so that only the new costs associated with establishing an assessment program are counted. When these costs involve reassignment of existing personnel, the full cost of replacement provides the basis for the estimate.

Constructing Tailored Institutional Cost Estimates

Because institutions vary widely in size, programs, and clientele, appropriate assessment programs will vary as well. A small, private, residential, liberal arts college will probably emphasize general education in its instructional mission to a degree not typical of a community college or a large research university. Consequently, it will appropriately concentrate its data gathering and analytical resources on assessing liberal learning outcomes. In contrast, a community college most likely will concentrate the design of its assessment program on job success and transfer to senior institutions. The clientele of the small liberal arts college will be much more conducive to administering tests and surveys in classroom settings than will be the case for the more dispersed community college population. As a result, methods for administering tests and surveys will vary considerably among types of institutions.

For illustrative purposes, we have constructed typical assessment programs for four types of institutions. They include: (1) a private liberal arts college with a traditional, residential student population of approximately

1,000 students; (2) a major public research university with a total student population of approximately 25,000 students (including 18,000 undergraduates); (3) a regional, comprehensive, public university with approximately 5,000 residential and commuter students; and (4) a mid-sized community college with an enrollment of approximately 15,000 students in occupational, transfer, and community service programs.

For each institution, we produced cost estimates as follows. First, based upon presumed instructional mission, we made a choice about which assessment dimensions should be emphasized. Second, we selected a typical array of instruments for each case, and estimated the direct costs for instrument procurement using published cost data for commercially available instruments, and common institutional experience for locally constructed instruments. Third, we chose a set of administration and analysis methods based upon expected student characteristics. Finally, we estimated coordination costs on the basis of the experience of existing data gathering and analysis investments in like institutions. In all four cases, we used actual data on costs incurred by similar institutions to support these typical programs. These data were provided by a total of eleven institutions with which we have worked closely on gathering assessment data and on using assessment results to improve program planning and decision making.

We constructed all four estimates by means of a specially-designed microcomputer template using the Lotus 1-2-3 Spreadsheet program. The template embodies available cost data on eight commercial test and survey instruments, as well as routines for estimating the costs of designing local test and survey instruments and of administering tests and surveys in classroom and mailed formats, and for estimating overhead costs associated with establishing an assessment office. The template contains on-line instructions for creating cost estimates.

Case 1—Private Liberal Arts College

Case 1 is a small private liberal arts college with a total enrollment of approximately 1,000 students. The student body is "traditional"; more than 95% attends full-time, and more than three quarters is in residence, living either in dormitories or in nearby private housing. The curriculum is also traditional, including a recently reinstituted general education core program and a typical list of undergraduate arts and sciences majors. There are no explicitly professional or pre-professional programs, although many students go on to professional or graduate training.

Assessment in this case is concentrated on the gain, or "value-added," of the total college experience, particularly in relation to its general education component. Because of the college's mission, the faculty have decided to administer the ACT-COMP Composite Examination to incoming freshmen and to graduating seniors. They have also opted to make maximum use of the COMP through a consulting visit each year in which ACT staff work with faculty in interpreting scores. The college has found that these visits are an important faculty development tool in addition to the information provided by the examination itself.

The college already participates in the CIRP freshman survey to a limited degree, and will supplement the sample to include the entire estimated freshman class (300 students). At the same time, interest in the involvement of currently enrolled students on campus led to a decision to administer the Pace College Student Experiences Questionnaire (CSEQ) to a selected sample of all students (150) each spring. Finally, the college conducts an alumni study every three years, covering the last three graduating classes. The college plans to develop its own survey but meanwhile is using the ACT-ESS Alumni Survey, which it supplements with 10 locally designed questions.

ACT-COMP testing occurs in classroom settings with dorm counselors serving as proctors. Each student receives an announcement of the test date and is provided with his or her own results after scoring. CIRP and CSEQ surveys are administered in class or through campus mail. The major survey effort is the alumni survey, but the small numbers of actual graduates each year do not entail a major cost. The response rate averages 75% for these surveys.

To coordinate the testing program, the college has appointed a junior faculty member in psychology as an assessment director at .35 FTE. She is assigned a 1/3 time secretary to handle announcements, record survey results, etc. Both personnel costs are shown in the estimate as the full replacement costs for these positions. As noted above, however, these will vary with the need for replacing these individuals in their current functions and with the current market costs of such replacement. Overhead costs are already absorbed by the office of the Dean of the Faculty to whom the assessment director reports.

Total estimated costs for Case 1 are documented in Table 1.

TABLE 1

Case 1—Private Liberal Arts College

Instrument Costs		
300	Freshman General Education Exams (ACT-COMP)	$4,500.00
150	Senior General Education Exams (ACT-COMP)	2,250.00
150	Senior Activity Inventories (ACT-COMP)	525.00
300	Freshman Surveys (CIRP)	415.00
150	Current Student Surveys (Pace CSEQ)	337.50
150	Alumni Surveys (ACT-ESS)	147.50
Administration Costs		
	In-Class Test Administration	
	Proctors, etc.	342.00
	Announcements, etc.	177.50
	Mailed Survey Costs (2 mailings)	193.62
Overhead/Analysis Costs		
	ACT Comp Consulting Visit (Fee + Travel)	1,375.00
	CIRP Data Analysis	150.00
	Testing/Measurement Specialist (.35 FTE)	9,625.00
	Secretary/Clerk (.35 FTE)	5,775.00
	Staff Benefits	3,388.00
TOTAL		$29,201.12

Case 2—Major Public Research University

The second case is a major public research university with a total enrollment of over 25,000 students, including about 18,000 undergraduates. Faculty make considerable research contributions to their own disciplines and concentrate much of their teaching energy on graduate instruction. Most introductory courses are lecture classes and are partially staffed by graduate teaching assistants. Most undergraduate students attend full-time, and about two-thirds are residential. Attrition rates are significant, but about 65% of entering students complete their degrees. Professional schools account for approximately 60% of undergraduate enrollment.

Because of its emphasis on professional and pre-professional study, much assessment effort has gone into testing in the major field. Graduates of about 10 programs per year are tested using available standardized test instruments. This year, 450 students are to be tested using a variety of GRE Field Examinations, and 360 students are to be tested using pre-professional examinations such as the National Teacher Examination (NTE) and the AICPA exam. In addition, the institution is evaluating general education using the ACT-COMP Objective Test in a test-retest format for freshmen and seniors. Like Case 1, the institution has budgeted for a consulting/faculty development visit in conjunction with the COMP.

To examine student life, the university intends to design its own survey, using faculty expertise. The survey will be administered to a stratified random sample of currently enrolled students in the spring. Because of the size of the campus and the characteristics of the sample, a mailed format will be used to administer the survey. Similar surveys in the past have obtained a 65% response rate.

To coordinate the testing program, the university has staffed an existing student research office with two new staff members, a testing specialist and a secretary. Existing senior staff in the testing office are also used in interpreting test results and in working with individual program faculties on improving curricula. Because many of the fields offered by the university are not now covered by an available standardized senior-level examination, testing center personnel are expected to work with program faculty to design local achievement tests as part of their budgeted assignment. Three such fields are scheduled for test construction this year in geology, archeology and food technology. Approximately $2,500 per test is budgeted for this activity to be paid to participating departments. Other test development costs are expected to be covered through department-level reallocation of faculty staff time already committed to curriculum review.

Table 2 presents total estimated costs for this case.

TABLE 2

Case 2—Major Public Research University

Instrument Costs		
2500	Freshman General Education Exams (ACT-COMP Objective Test)	$15,000.00
1700	Senior General Education Exams (ACT-COMP Objective Test)	10,200.00
450	Senior Field Exams (GRE)	13,050.00
360	Senior Field Exams	7,500.00
	Development Cost for 3 Local Field Examinations (Professional and Pre-Professional)	9,270.00
	Development Cost for Student Survey	5,200.00
2025	Surveys (Production and Scoring Cost)	518.75
Administration Costs		
	In-Class Test Administration	
	Proctors, etc.	1,826.00
	Announcements, etc.	2,077.00
	Mailed Survey Costs (2 mailings)	1,957.00
Overhead/Analysis Costs		
	ACT Comp Consulting Visit	1,375.00
	ACT-COMP Data Tape	20.00
	Testing/Measurement Specialist (1 FTE)	27,500.00
	Secretary	16,500.00
	Staff Benefits	9,680.00
	Office Expenses	8,400.00
TOTAL		$130,073.75

Case 3—Regional Comprehensive University

Case 3 is a public regional comprehensive university enrolling approximately 5,500 students, including 4,500 undergraduates. Like many of its type, the university is a former teachers' college which became a comprehensive university in the early 1970's. In addition to liberal arts disciplines, the university now offers a range of professional subjects through the master's level. These are dominated by education and business, which together enroll about half of the student body. Forty percent of the undergraduate students attend part-time and about two-thirds commute. The university does not currently commit significant resources to academic administration and support, and is proud of its tradition of "low overhead."

As in Case 2, the university seeks to insure that graduating seniors have received adequate training in the major field. Therefore, it has chosen to administer standardized senior examinations annually to the graduates

TABLE 3

Case 3—Regional Comprehensive University

Instrument Costs		
300	Freshman Tests (ACT Assessment: 900 assumed to have scores on entrance)	$2,100.00
800	Sophomore Tests (ACT Assessment)	5,600.00
1200	Freshman Interest Inventories (ACT Assessment)	3,000.00
80	Senior Field Exams (GRE)	2,320.00
40	Senior Field Exams (Professional and Pre-Professional)	940.00
1200	Entering Student Surveys (ACT-ESS)	240.00
350	Non-Returning Student Surveys (ACT-ESS)	70.00
650	Alumni Surveys (ACT-ESS)	130.00
	Scoring for 2200 ACT-ESS Instruments	1,040.00
Administration Costs		
	In-Class Test/Survey Administration	
	Proctors, etc.	375.00
	Announcements, etc.	580.00
	Mailed Survey Administration (2 mailings)	1,378.00
Overhead/Analysis Costs		
	ACT-ESS Tape/Reports	270.00
	Testing/Measurement Specialist (.35 FTE)	9,625.00
	Staff Benefits	2,118.00
	Work Study Students	1,750.00
	Office Expenses	1,250.00
TOTAL		$32,786.00

of a fifth of its departments on a rotating basis. This year approximately 120 graduating seniors in fourteen fields will be tested using a variety of instruments. GRE or pre-professional examinations are used where possible. All but three fields in which degrees are offered by the university currently are covered by an existing standardized examination.

The university is also committed to building basic skills, but its emphasis on general education is insufficient to justify the expense of an instrument such as ACT-COMP. Therefore, the faculty has decided to examine "value-added" by using the ACT assessment administered to entering freshmen and at the end of the sophomore year. Of the approximately 1,200 new freshmen each year, about 300 must be given the ACT assessment at university expense. All 800 sophomores are subsequently tested at university expense.

Finally, the university has elected to use a relatively low-cost, standardized survey system—the ACT Entering Student Survey—to investigate student opinion and post-graduate success. All entering students are surveyed using the ACT-ESS, and all graduates are surveyed a year after graduation. Finally, a sample of withdrawing students is followed up every other year with the ACT Withdrawing Student Survey. Entering student questionnaires are administered at freshman orientation. Other surveys are administered by mail. All scoring is done by ACT, although the university purchases extra reports and tapes for local analysis. Few local analyses of these data, however, have actually been conducted.

To coordinate testing, the university grants 1/3 release time to a faculty member in sociology. As before, the full cost of replacement is provided here, although replacement costs using part-time instructors have in practice been less. Work-study students are used to support the survey effort, and graduate students in education are used as test proctors for sophomore and senior examinations.

Costs for Case 3 are itemized in Table 3.

Case 4—Mid-Sized Community College

Case 4 is a community college located in a suburb of a major city, enrolling approximately 15,000 students each term. Enrollment consists of about 3,500 baccalaureate transfer students and 5,000 students in various occupational and certificate programs, with the balance enrolled for one or more single courses. About half of the students in baccalaureate transfer and occupational programs attend full-time, many of them at night. All other students are part-time attenders. All students commute to the campus from within a 30-mile radius.

The primary emphasis of assessment at the college has been properly placed upon student follow-up and the assessment of educational goals. All entering program students are surveyed at registration using the NCHEMS/College Board Student Outcomes Information Service (SOIS) Entering Student Questionnaire. In addition, each year those completing a program and those withdrawing from programs are surveyed by mail using SOIS instruments. These surveys achieve approximately 70% and 45% response rates, respectively. Local questions are added to all SOIS questionnaires, and the Institutional Research office conducts analyses which link common questions on the three instruments in order to obtain a composite picture of student reactions to the college experience. All SOIS scoring is done by the College Board, although the college plans to develop its own computer programs to facilitate a more detailed analysis of these instruments.

This year, reacting to statewide concerns about the quality of basic skills education, the college plans to administer the College Level Examination Program general exams in writing and quantitative skills. This will be an expensive effort and is being undertaken somewhat reluctantly. Despite the fact that it was not designed for curriculum evaluation, the CLEP was chosen by faculty as being the most appropriate available instrument to measure general lower-division competence in these areas. It will be given to a sample of 750 second-year program students.

The student follow-up effort was begun several years ago in response to federal Vocational Education Data System (VEDS) requirements. This system mandated student follow-up surveys for graduates of all federally-funded occupational programs. To meet the demand, the college created a half-time data analyst position in the Office of Institutional Research. As assessment has expanded, the responsibility for conducting all studies has remained with Institutional Research; an additional half-time position will be added for survey coordination and to help with administering the CLEP.

Total estimated costs for assessment at Case 4 are presented in Table 4.

TABLE 4

Case 4—Mid-Sized Community College

Instrument Costs	
750 Sophomore General Skills Exams (CLEP General)	$19,500.00
4500 Entering Student Surveys (SOIS)	675.00
1500 Former Student Surveys (SOIS)	225.00
1250 Graduate Follow-Up Surveys (SOIS)	187.50
Scoring for 7250 SOIS Instruments	3,490.00
Administration Costs	
In-Class Test Administration	
Proctors, etc.	325.00
Announcements, etc.	225.00
Mailed Survey Administration (2 mailings)	3,693.00
Overhead/Analysis Costs	
Tapes/Reports of SOIS Surveys	150.00
Student Survey Coordinator (.5 FTE)	10,750.00
Staff Benefits	2,365.00
Office Expenses	1,500.00
TOTAL	$43,085.50

To check the validity of each of these cost estimates, we obtained actual cost data from a total of eleven institutions covering all four of our "case" categories. Total costs for assessment at each of these institutions, of course, vary somewhat from our constructed estimates and from one another. This variance occurs because each institution measures a somewhat different set of outcomes dimensions, and also because the manner in which actual costs are counted and reported are different between institutions. For reasons of confidentiality, we do not report these actual costs. However, when adjusting for total enrollment, none differs by more than 15% from our constructed estimates.

Some Concluding Points

Each of the cases presented above represents a distinctive match between institutional mission and characteristics on the one hand, and a particular choice of assessment instruments and methods on the other. Any cost estimate must be similarly tailored to fit a particular situation. In conclusion, institutions considering implementing a comprehensive assessment program and examining the cost consequences should consider the following points:

- Making full use of *existing* information about student learning and development can considerably reduce anticipated costs of assessment. As emerging institutional experience makes clear, colleges and universities generally collect considerable information about students; but this information is rarely centrally available. Indeed, no single person or office at the institution may know the full range of what is available. Many individual units may collect data for different purposes. For example, individual departments may collect follow-up information on their own graduates, student service offices may conduct surveys of currently enrolled students, and testing offices may administer a variety of standardized tests. A first step in constructing an assessment program is often simply to inventory such data (Ewell, 1982).

- Developing an explicit assessment program may reduce cost by focusing analytical and data collection resources and avoiding duplication. Emerging institutional experience has also shown that gathering data on student outcomes can often be inefficient due to its dispersal throughout the institution. Different units develop their own assessment instruments independently, and incur costs in doing so. Furthermore, many studies are one-shot—designed to answer a particular question or to address a particular, temporary crisis. When the question is answered or the crisis passed, data gathering ceases, only to be begun from scratch when the next question arises. Central coordination of assessment can avoid such hidden costs, and may consequently involve fewer new resources than initially anticipated.
- Assessment programs using multiple data collecting methods may similarly reduce costs by providing mutually reinforcing information. Cognitive testing, for example, is expensive compared to other forms of outcomes data gathering. While there is no substitute for testing to answer questions concerning student learning in general education or in the major field, much can be learned by supplementing cognitive tests with less expensive kinds of data collection—for example, student surveys containing self-assessments of growth. If survey information can confirm the results of cognitive tests in the aggregate, expensive testing for purposes of program evaluation may then be undertaken using only small but carefully selected samples of students.
- Careful tailoring of data collection to fit instructional mission can limit cost. A major potential problem with assessment, as mentioned above, is the implicit assumption that it will "measure everything that moves." Paying close attention to priority instructional and curricular issues in designing an assessment program involves making appropriate choices about what to measure and how to measure it. Each of the cases we constructed, for example, places the primary weight of assessment upon a particular dimension that matches the institution's unique curriculum and mission. Each could have been quite different, and considerably more expensive, if limiting choices had not been exercised.

A final point is that the costs of assessment are in themselves of little importance without knowing the benefits. Many of the 22 institutions involved in the NCHEMS/Kellogg Student Outcomes Project in 1982-85, for example, found that the long-term benefits of assessment information included increases in student recruitment and retention (Ewell 1984). In the long term, such benefits can involve fiscal, as well as strictly educational, rewards. As a result, any assessment program is properly seen as not simply a cost to be incurred, but as an investment in the institution's future: an investment which should be judged in the light of the return that it may bring.

References

American College Testing Program. *Evaluation Survey Service*. Iowa City, Iowa: ACT, 1985.

Astin, A. W., K. C. Green, W. S. Korn, and M. J. Maier. *The American Freshman: National Norms for Fall 1984*. Los Angeles: Higher Education Research Institute, 1984.

Chancellor's Office, California Community Colleges. *Item Bank: Student and College Outcomes*. Improving Community College Evaluation and Planning Project. Sacramento, Fall 1984.

Dillman, Don A. *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley and Sons, 1982.

Ewell, Peter T. *The Self-Regarding Institution: Information for Excellence*. Boulder, CO: National Center for Higher Education Management Systems, 1984.

Ewell, Peter T. *Student-Outcomes Questionnaires: An Implementation Handbook*, 2nd Edition. Boulder, CO: NCHEMS, 1983.

Ewell, Peter T. *Information on Student Outcomes: How to Get It and How to Use It*. Boulder, CO: NCHEMS, 1982.

Forrest, Aubrey. *Increasing Student Competence and Persistence: The Best Case for General Education*. Iowa City, Iowa: ACT, 1982.

Higher Education Research Institute. *Composite Comparative Data College Student Experiences Questionnaire 1979-1982*. Los Angeles: HERI, 1982.

Lenning, Oscar T. *The Outcomes Structure: An Overview and Procedures for Applying It in Postsecondary Education Institutions*. Boulder, CO: NCHEMS, 1977.

Mentkowski, Marcia and Austin Doherty. *Careering After College: Establishing the Validity of Abilities Learned in College for Later Careering and Professional Performance*. Final Report to the National Institute for Education. Milwaukee, WI: Alverno Productions, 1983 (ED 239 556-239 556).

McKenna, Barbara. *Surveying Your Alumni: Guidelines and 22 Sample Questionnaires*. Washington, DC: Council for Advancement and Support of Education, 1983.

National Institute of Education, Study Group on the Conditions of Excellence in American Higher Education. *Involvement in Learning: Realizing the Potential of American Higher Education*. Washington, DC: U.S. Government Printing Office, 1984.

Pace, C. Robert and Associates. *Higher Education Measurement and Evaluation Kit*. Los Angeles: Laboratory for Research on Higher Education, UCLA, 1975.

Assessment in Higher Education: To Serve the Learner

by Georgine Loacker, Lucy Cromwell, and Kathleen O'Brien[1]

Assessment seems to be loitering expectantly in the corridors of higher education, thereby reinforcing the hope that it will soon enter the classroom to serve the learner. Recent national reports on higher education encourage assessment. Administrators call for it. Researchers see it as a potential instrument for prediction and evaluation. Legislators look to it for assurance of accountability. But many of these intents overlook the power of assessment for teaching and learning. So that teachers might take a more serious look at assessment, we propose in this paper to set it at the heart of learning and to clarify it as a major strategy to be used by both instructor and student.

Though the word *assessment* did not emerge from classroom or campus, it derives from an idea important to educators—*that of sitting down beside or together* (from late Latin *ad + sedere*). In the seventeenth century an assessor was "one who sits beside" or "who shares another's position." Early uses of the word focused primarily on determining the worth or value of something in monetary terms, but underlying those uses was the idea of expert judgment made on the basis of careful observation. "Assessment" was thus a word destined for the tongues of educators—whether humanists or scientists.

Definitions and Assumptions

Assessment, as we use it throughout these pages, is a multidimensional process of judging the individual in action. Embedded in this definition are assumptions about learning that emphasize active development of the learner.[2]

Assumptions. One assumption is that *learning involves making an action out of knowledge—using knowledge of think, judge, decide, discover, interact, and create*. We contend that acquiring or storing knowledge is not enough. Unless one carries knowledge into acts of application, generalization, and experimentation, one's learning is incomplete.

Another assumption is that an *educator's best means of judging how well a learner has developed expected abilities is to look at corresponding behavior*—thinking behavior, writing behavior, inquiry behavior, or appreciating behavior, for instance. We presuppose a link between behavior and cognitive and affective processes. Because human behavior is purposeful, educators can find out more about a learner's problem solving ability by observing that person actually solving a problem and clarifying reasons and processes than by confirming a "correct" solution he or she has selected from a set of alternatives.

A third assumption is that *learning increases, even in its serendipitous aspects, when learners have a sense of what they are setting out to learn, a statement of explicit standards they must meet, and a way of seeing what they have learned*. When students of science, for example, are told that they will have to go beyond reading their text, listening to their teacher, and replicating lab experiments—that they will have to raise their own questions and test their own hypotheses—they are more apt to learn to do all of the above more meaningfully and effectively. Out of that success they then develop confidence that enables them to recognize unsought-for insights when they come upon them.

We contend that such awareness of expectations and standards enhances learning because it places in a person's hands the means of collaborating in his or her own learning and gradually taking control of one's own learning process. Within that context, learners recognize that their question, "How am I doing?" is taken seriously. They also begin to see an important implication of that question: that further learning builds on, and develops from, where each learner is at any given point. Therefore, that question becomes the occasion for doing better when everyone responsible for learning—teacher as well as student—receives as complete an answer as possible. Assessment aims for such an answer.

What does it mean to aim at an increasingly complete answer to the question of how a person is doing? One can get some insight into the question by considering what testing traditionally tells us about someone, in contrast to what assessment tells us.

Assessment in Contrast to Testing and Measurement. Testing, as it is frequently practiced, can tell us how much and what kind of knowledge someone possesses, whereas assessment provides a basis for inferring what that person can *do* with that knowledge. Much testing carefully limits what we can know about a person to a set of written or marked answers. Assessment aims to elicit a demonstration of the nature, extent, and quality of his or her ability in action.

When we narrow testing to measurement, it answers the question "How am I doing?" with a quantitative response that says, "You did a certain percent of what was asked on a given occasion" or "You did as well as a certain percent of all those who tried or might try to do the same." Assessment answers the question with a descriptive account of precisely what the individual person has done on a given occasion. By judging a person's performance against pre-set, agreed upon, and public criteria, assessment aims to make the performance meaningful so that he or she can build future performance on the basis of understanding.[3]

Assessment and Evaluation. Emphasis on the progress of the individual learner also distinguishes assessment from program evaluation. Evaluation looks for elements that can be combined and compared in order to draw conclusions about groups of students, with a view to making judgments about the general direction of a course, program, or curriculum. Assessment looks for distinguishing elements in a person's performance and relies on varying contexts to assure that as much complexity of a person's ability is elicited as possible.

Our definition of *assessment* is shaped by its power to serve the learner; it means eliciting samples of varied expressions of an ability, judging those samples against identified criteria for performance, and providing as full a picture as possible of that ability as possessed by that learner. Assessment as learning weaves together several strands of a long history of meaning that have developed separately.

History

The practical history of assessment in business and government is essentially the history of the Assessment Center Method. And it is—at least until recently—the history of improved selection and screening rather than of development and learning. In the 1930's, when it began in England and Germany, assessment provided a new, behaviorally oriented means of selecting military officers. In the 1940's, with researchers from the Harvard Psychological Clinic adapting and further developing assessment, the United States Office of Strategic Services used it to select American intelligence agents. In the 1950's, led by AT&T, business and non-military government departments contributed to the extensive growth of assessment centers by using them to select managers. More recently, business and government have begun to show interest in using the assessment center method for development.[4]

Assessment centers have established several concepts from which education can benefit. As it characterizes the assessment center method, assessment involves behavioral descriptors to develop a rich picture of an individual's ability, uses multiple techniques for judging performance, and refines assessor judgment through articulation of more explicit evidence.

Assessment in Education and Psychology

In education, the term assessment is used in broad, varying senses. Its most frequent use is as a synonym for program evaluation. However, in all of the above contexts, as well as in the context of clinical psychology, the word assessment emerges often in contrast to *testing* and connotes a concern with broader educational outcomes than knowledge.

For nearly 50 years, psychologists have tended to use the term in relation to broad and multiple abilities. They thus added to the connotation of the word the concept of abilities, taken not as static traits but as processes, and thus changeable and directable. Even as the denotation of the word has become more general and diffuse, the connotations have emphasized multiple performances and breadth of abilities. There are now several indications that the educational world is adopting these connotative meanings and is receptive to the idea of assessment as we define it in this paper.[5] National reports have asked educators to be accountable by reinstating the learner at the center of higher education. The phenomenon of increasing enrollment of adults in college is reminding educators that learning is developmental and continues throughout the life span. It also suggests that one cannot evaluate experience as prior learning unless one defines learning in terms of developed abilities or significant expected outcomes. Emphasis on experiential learning—doing what one knows—has surfaced as an important component of the learning cycle of an individual. Finally, current questions about the usefulness of intelligence measures and of standardized tests such as the SAT also focus attention on the need to develop other approaches to assess an individual's ability and potential to learn.

Conceptual Elements of Assessment

Every teacher has had the experience of hearing some version of the young Helen Keller's cry of "Water," the experience of discovering a student's sudden illumination or success. And once having heard it, who does not wish to find a way of making it more frequent, more developmental, and more characteristic of every student? Teachers need to find ways to build on, and expand, moments of learning for all students, rather than merely rewarding them.

Assessment becomes a meaningful way to expand learning when one defines it to include a set of key elements that make it a learning experience. It provides a way of refocusing education on individual learners instead of using a wide lens on an indistinguishable mass from which we can infer only general patterns. Since students are grouped in courses, the idea of using assessment as a camera that takes individual portraits instead of group pictures requires explanation. It is essential that a dynamic, cumulative, and composite picture of a student's abilities be made visible to everyone responsible for the student's learning—including the student.

To create such a picture, assessment needs to be defined to include multidimensional sampling of student's abilities in action, observation and judgment of those samples on the basis of explicit criteria, and structured feedback administered sequentially in relation to a learner's development. Each of these elements in turn must contribute to the growth of students' abilities to assess themselves.

Sampling Student Performance

Observing a student in action brings us as close to an individual's ability as we can get. Because we cannot observe all of a person's expressions of a given ability, we take intermittent samples. Given the complexity of the human being, there will always be a distance between behavioral data and the ability itself. Even a very precise image of exactly how a detective has gone about solving a mystery offers a very limited view of his or her full detecting powers. Sampling is at least a start toward developing a picture of an ability in operation.

Without a behavioral sample, instructors can look at a set of selected answers and judge whether a person was able to recognize given facts or concepts. They can look at a description of what a person says that he or she knows about something and would do with the knowledge. But those instructors can only assume and hope that the knowledge can translate into effective action.

With a behavioral sample, we can at least see that a person did or did not do something in a given context. Therefore, they can say that someone *can* do it—at least in such a context—whether or not he or she *will* do it again. For the persons assessed, sampling provides a picture that enables them to look from the outside at their own ability in action, to supplement their inside view.

To assess, therefore, requires that we sample students' behavior. We need to sample their writing to judge whether they can write. We need to sample their synthesizing to judge whether and how they have put together the facts they have learned. We need to sample students' work in groups to judge whether they can think and work collaboratively.

Multidimensionality. In order to elicit enough dimensions of behavior for fair judgment, sampling needs to be multiple and varied.

Multidimensionality provides a means of addressing some of the questions that single sampling raises: How do we know a sample is representative? Is the person having an unusually good or bad day? Can and will the person repeat the performance under different circumstances? The only way we can begin to form an answer is to gather enough samples to enable a pattern to emerge. Perhaps it turns out that an unusually good day—or a bad one—seems to be representative, or more likely, that either or both are occasional occurrences. In either case, it is important to be as precise as possible in discerning the elements that constitute each performance so that the assessee's knowledge and experience of them can refine strength and transform weaknesses.

How varied need the samples be to suggest the complexity of an ability? Such variables as a written or oral mode, a static or dynamic object of analysis, a solitary or collaborative responsibility for accomplishing a task, all evoke different dimensions of an ability. Being able to analyze written data at one's desk, for example, does not mean being able to analyze data as it occurs before one's eyes in a group situation. Nor can good writers always organize their thoughts as well when speaking.

The reasonable response to the question of varying contexts, therefore, seems to be to vary them according to the ordinary shifts of life situations, such as the purpose, the nature and number of people involved, or the amount of information available. If careful feedback is provided, each shift in context can assist learners to refine their understanding of an ability and how they exercise it. The effect of varying context is twofold; it reinforces the general core skills involved, and it reveals unique skills elicited by each situation.

The success of the "writing across the curriculum" movement dramatizes the new understanding of educators that effective writing as a life ability needs to be practiced and assessed in a variety of disciplines, in fact, in every discipline studied. If writing across the curriculum makes good academic sense, why not the assessing of other abilities across the curriculum?

How many samples are necessary to provide a full picture of a persons's ability? From the tens of thousands of hours of a student's academic career, we can select but a few for careful observation. Through these few hours we need to get as full a picture of the student's ability as possible and thus create an increasingly secure basis for judgment. We can do that by using the other major components of assessment—observation, judgment, explicit criteria, feedback, and self-assessment—with a view to shaping a process that makes single assessments complementary and cumulative. Such a process serves the learner by clarifying a pattern that shows the unique highlights and shadows, the fullnesses and gaps of a picture that takes shape gradually with each new line affecting the direction of the next one. For example, by looking at successive samples of their writing, with feedback that focuses on agreed-upon criteria, learners can better understand their ability on a developmental basis. They can also see how varied purposes and audiences elicit unique characteristics of their writing and sometimes heighten their strengths or depress their weaknesses.

Observation

Assessment calls forth from teachers their keenest powers of observation. It depends on their ability to set aside tendencies to quantify and rank, or to eliminate, possible alternatives. An effective assessor looks at what is happening behaviorally—at a student drawing conclusions, for example, whether at a podium or in a paper. Such observation involves attention to parts in precise relationship to each other and to a whole, including emphasis and proportion. It involves adopting an open framework to preclude any tendency one might have to look only for error or to be biased by a single expectation.

Such a framework is built on the criteria of performance that one gradually develops from experience—by reflecting on good performances and attempting to articulate the basis for one's judgment. That framework represents an increasingly expansive understanding of an ability. One important aspect of that understanding is recognizing the limits of the framework; as an organization of criteria of performance, it never fully describes the ability. It allows, however, a range of varied expressions and styles that contribute to the overall effectiveness of student performance and to the uniqueness of individual ability. In presenting conclusions from experience, for example, some learners begin with detailed descriptions of their experience and then abstract general principles. Other learners initially seize upon general principles and then accumulate evidence to support them. The effectiveness of the former lies in the ability to engage readers' or listeners' minds with the immediate before leading them to the abstract. The effectiveness of the latter lies in the ability to set forth points with clarity and gradually convince with supporting evidence.

Externality. To observe a developing ability in action requires a perspective outside the direct interactive teacher/learner process. This external perspective might come from criteria established throughout the department, or from assessing done by someone other than the student's course instructor or from college-wide assessments that call forth the integration of content and/or skills from more than one course.

Even in regular classroom assessments, teachers need to establish a measure of distance to assure that a new judgment is made on the basis of criteria applied to a specific situation, rather than one limited to a series of evaluations already recorded. Otherwise they have no guarantee that their observation and judgment make a fresh addition to their accumulated understanding of a student's ability.

Self-assessment is even more of a challenge. The struggle to stand outside of one's own performance is essentially what makes learning to assess oneself so long and complex a process. Practice—in looking at records of one's own performance and in general refining of one's ability to observe and judge according to criteria—makes self-assessment more attainable.

Judgment and Explicit Criteria

The experience of faculty as expert judges of student ability is an important reason for placing them at the center of any educational assessment process. Even faculty who have never verbalized their standards and who might use a norm-referenced framework to report their judgments, work from an implicit understanding of what they expect in student performance. Assessment requires them to articulate that understanding in explicit and public statements of criteria of performance. By doing so, faculty refine their own understanding of expected abilities, clarify for their colleagues the basis of their judgment, and enable students to understand what performance is required.

Explicit criteria provide a major means of getting a picture of an ability, for they serve as indicators of that ability as seen in performance. Thus they are one of the components of assessment that distinguish it as learning. The picture sketched by criteria should be sufficient to enable the assessor to judge the presence of an ability. It also needs to be clear enough for the beginning learner to imagine a performance that would match the criteria.

Criteria, as we define them, are standards external to the object of judgment, used to identify those characteristics of the object that indicate its worth. They are articulated by faculty acting as expert judges of performance, in a process of clarifying the holistic judgments they have made throughout their experience on the basis of their skill in a given field.

Because assessment in an educational setting must deal with multidimensional abilities, we do not suggest a precise formula for stating criteria. Some criteria can be easily applicable with little judgment—"Follow the APA manual of style," for instance. Others necessarily require greater use of judgment and clarification of specific aspects of a situation—"Defends own position adequately" or "Shows quality of workmanship."

When students perceive performance criteria to be learning objectives, when students discover, for example, that they do not meet the criterion of "appropriate use of linguistic conventions" or "adequate development of ideas," then assessment becomes learning itself.

Research on Criteria

Perhaps the most persistent question about explicit criteria of performance is how specific they should be. Our research at Alverno College suggests that the context of the developmental level of the student is a significant determinant of the degree of specificity.

Beginning students. We find that students at the start need very explicit criteria. They are trying to figure out what they're supposed to do and, in effect, they use the criteria as a recipe or set of directions to plot a performance. Initial results from the longitudinal study conducted by the Alverno Office of Research and Evaluation support this impression. They indicate that students begin with the perception that criteria are directions for what and how much to learn and that competencies are directions for what to do. While these students see highly detailed directions as "picky," they see broader directions as "vague."

After a semester or two, students begin to cluster the criteria they had formerly seen as discrete steps or directions and recognize that the criteria are related, that they come together to define an ability. Students begin to realize, for example, that making inferences and supporting them with data are not complete steps in themselves, but are part of the ability to think critically. Thus, students gradually begin to see more complex abilities contributing to effectiveness in their performance.

Advanced students. At more advanced stages of education, students have begun to develop their own understanding of an ability, and specified criteria serve to supplement what learners have internalized or to remind them what they have not yet internalized. Having developed a range of abilities to call on in varied situations, advanced students should be able, given a context, to infer the kind of performance elicited, call upon the required abilities, and infer criteria of performance. According to Alverno research, the most advanced students begin to internalize the need for criteria; they see criteria as part of self-assessment and use them to guide their learning. At this level, criteria can be stated holistically. For example, a student might be told that "thorough analysis" is a criterion for her performance. Both student and teacher understand that "thorough analysis" means applying a framework, identifying elements and relationships, supporting inferences with evidence, and so on.

Interpreted thus, criteria of performance constitute the primary tool of the assessor—especially if the assessees are expected to learn from the experience and to become assessors of their own performance.

Sequential Administration

Assessment can serve learners best when they can carry a developing picture of their abilities from one assessment situation to the next. Students can make some of those connections for themselves when faculty identify what is to be assessed, what criteria will be used to judge it, and how well it has been done. But once learners know how well they have done in one assessment situation, and have an idea of how they might improve, they need opportunities to demonstrate their improvement.

Within a course, therefore, formative assessments need to build on each other in a way that is clear to the student. And summative assessments need to build on the formative and on each other. In fact, if an institution expects of students some outcomes that transcend courses—and all colleges do in relation to both the major and general education—then faculty must provide sequenced, external assessments to give students opportunities to integrate the knowledge and abilities they have demonstrated in discrete courses. In effect, in order to address the student's ongoing, overall development as a learner, faculty must extend assessment across the curriculum, and that assessment must be developmental as well as reinforcing.

Feedback

For assessment to be learning, feedback is critical. Feedback offers the teachable moment, the opportunity for change. It takes the elements of assessment discussed thus far and turns them into learning. It can be seen as both a resource and an event. As a resource, it is information provided by the assessor, and in some cases by the assessment itself, which presents a profile of how the learner in action meets criteria of effectiveness. As an event, feedback is the time when the learner and assessor "sit down beside each other" and direct their attention to the strengths and weaknesses of the learner's performance.

"Sitting down" can mean that the student and faculty member have a face-to-face interaction or that a course instructor gives feedback to the entire class and to small groups within the class. It can also take the form of a well-worded sentence written from the faculty member to the learner. Whatever the form, feedback interprets performance as judged by criteria, thus extending the picture of a student's developed ability. It makes this picture available and revealing to both partners in the assessment process.

Feedback at its best is an opportunity to learn. It goes beyond indicators of rank in class or percentage of items correct to describe uniqueness, reveal strengths, and illuminate the basis of weaker aspects of the learner's performance. It suggests where to aim to develop an ability. By reinforcing the learner's understanding of what he or she knows, it motivates further development. In this latter sense, the moment of feedback is also a time to redirect efforts and make plans to practice nuances of the ability being developed.

Research on Feedback

Good teachers know that to be effective, feedback should be timely, informative, explicit, focused on what can be changed, and generally positive in nature. Still, they might ask how explicit feedback should be. Should negative feedback be given, and if so, when? What level and amount of information constitutes optimal knowledge of results?

Experience in providing feedback at Alverno suggests that one way to deal with these questions is to study the developmental stages of learners in relation to their use of feedback. While knowledge of "stages" is still incomplete, what the faculty do, know, and report here has been helpful in working with students.

Beginning students. Beginning students prefer specific, concrete feedback. They focus on aspects of their performance as if these aspects were isolated and unrelated elements. Feedback that is positive, specific, and concrete helps at this stage, but is most effective if it assists the learner to see the relationships among the discrete elements of performance.

Another characteristic of beginning students (whether they are first-year students or beginning a new course of studies sometime later) is that they often let emotional responses hinder their insight. The instructional strategy should provide as much positive, specific feedback as possible in earlier assessments. For less successful elements of performance, instructors should provide feedback that points out why the student ran into difficulty and what concrete steps can be taken in order to improve. Care in these matters is especially important with students for whom knowledge of multiple weaknesses might tend to be overwhelming.

For instance, in a first-semester humanities course at Alverno, students who have written essays on the pros or cons of an aesthetic issue receive feedback on their analytic and writing abilities that concentrates on the positive. Faculty point out how the discussion takes account of the selected audience or where the writer offered clear relationships among the key arguments. However, they also point out at least one area that needs further development, as in the following example:

> You show awareness of the author's use of symbolism. Where I think you could improve is in reflecting on the meaning and significance of those symbols. One of the characteristics of a symbol is that it points to some larger idea. You need to be more explicit in identifying those broader areas.

Then, in order to assure further learning, an instructor might ask the student to review samples of the work of previous students who had effectively clarified the significance of specific symbols. These samples might be in a reserve file in the department or library, or they might be called up on a personal computer. Whatever the mode, such feedback challenges learners to move beyond their present ability and, by exposing them to a range of peer examples, gives them some idea of how to do it.

Advanced students. As students develop the ability to use feedback as new learning, they take a more objective stance toward their own behavior. They seek out evaluation of their work. They want feedback that helps sort out patterns and relationships among varied abilities and disciplinary contexts. Consequently, feedback to advanced-level students should place less emphasis on elements effectively demonstrated, and more emphasis on the learner's performance in relation to past work and to the nuances of the underlying ability.[6]

One strategy Alverno faculty have found effective for advanced students is to use the expected outcomes of the major as an organizing principle for feedback. For instance, history faculty have identified three major outcomes that each graduating student must demonstrate. One of these is the *ability to articulate, integrate, and employ methods of history to create a coherent understanding of her own and other cultural heritages.* Work submitted during the senior history seminar is assessed in light of this ability, and feedback to students indicates in what way and to what degree each student is demonstrating it. Consequently, feedback to advanced students aims to sketch an increasingly holistic profile of the learner as history major.

Self-Assessment

The ability to appraise one's own performance is not an automatic culmination of the learning process. To develop autonomy as learners, students must gradually try out strategies for achieving distance from their performance and applying criteria to it. Therefore, the ability to self-assess should be an essential component of the assessment process and an important part of each individual assessment.

Assisting learners to develop the ability to self-assess is a multi-dimensional process. It means teaching them to observe themselves in action. It requires students to develop the habit of asking what these observations mean about their own behavior and the underlying array of expectations, knowledges, and abilities that these behaviors represent. It asks students to make judgments about the effectiveness of their behavior in reference to a set of standards or criteria rather than making comparisons to the work of peers. Finally, developing complex self-assessment ability involves learners in finding more effective, yet distinct, models of performance that can serve as behavioral alternatives for future development.

Developing an Assessment: Guidelines for Faculty

How does one go about developing an assessment of the kind that we have been describing? And who is the "one" to develop it? How can one do so, particularly in light of the specific content of a course, discipline, or general education program, the particular level of student to be assessed, and the creative intelligence of a teacher? The process of assessment design is complex, as complex as the situation with which a faculty member deals whenever designing a learning experience or system. In order to make the process accessible, therefore, we will deal with it in an inductive fashion, working through the elements as any teacher might, and translate them into a design for developing assessments.

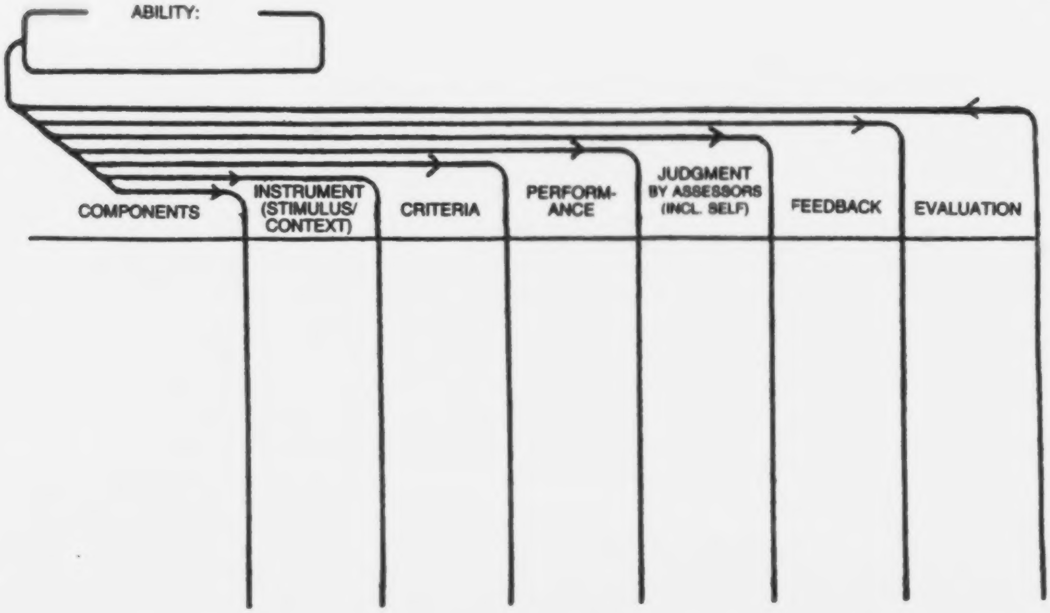
Who is the "one" who develops assessment? In defining *assessment* as an educational process we have stressed that it includes not only a specific evaluative event, but also the ongoing relationship between teacher and student and the even more cumulative sense of a student's overall development across the curriculum. Assessment is thus a responsibility shared by individual teachers and a college or university as a whole, and therefore we will present the design of assessment from two perspectives—first, that of the individual faculty member, then that of the larger curriculum.

Designing Individual Classroom Assessments

Let us imagine any teacher ready to design an assessment, and thinking aloud: I take as a basic working assumption that my aim is to *sample* my students' abilities and to provide *multiple opportunities* for that sampling. The method that I follow is not a rigid series of steps, but a logical pattern in relation to the elements of assessment. I begin by determining the outcome I expect—the ability I want my students to demonstrate. At some point, I have to determine a stimulus and context, and designs for feedback and self-assessment. Beyond any individual assessment, I consider how each assessment experience relates to the ongoing development of the student, especially in relation to other assessments in my course.

Each assessment that I design is part of a larger pattern. I remember what my students have demonstrated in the past; I anticipate what they will be learning in the future. My process, then, includes attention to the development of the student as an individual learner as well as to that student's participation as a member of my particular class. When I design a specific assessment, I try to insure attention to these several considerations.

At Alverno, a generalized model describes the flow of the process, assuring the inclusion of crucial elements and feeding back into an evaluation of each aspect:


ABILITY:
COMPONENTS
INSTRUMENT (STIMULUS/ CONTEXT)
CRITERIA
PERFORM- ANCE
JUDGMENT BY ASSESSORS (INCL. SELF)
FEEDBACK
EVALUATION

I may not always work with these elements in the same order, but I include them all each time I design an assessment. In use, the model is spread out and rearranged. For my design I can begin with any of the elements discussed below, but in doing so, I inevitably set up a series of connections. In the discussion that follows, my starting point is determining the ability to be demonstrated, but wherever I begin, connections between elements lead to the other steps in the process.

1. DETERMINE A SPECIFIC ABILITY OR EXPECTED OUTCOME

A major assumption underlying assessment, as distinct from traditional testing, is that learning—and by extension, assessment of that learning—should be designed to foster the growth of student abilities in significant areas beyond the acquisition of knowledge. As I work with students, I teach and assess a variety of abilities ranging from interpreting data to analyzing constructs for perceivable organizing principles. It is these that I must identify, integrating them with the content of my discipline. In designing an assessment, I examine the overall goals of my course to determine which of them, and with what degree of complexity, I can assess at a given time.

In an introductory fiction course, for example, one of my course goals is for a student to "show understanding of the way in which readers make meaning in literature by analyzing literary elements in relationship to each other." Such a goal embodies an attempt to focus on the content of my discipline as more than the specific "texts" to be studied—not only as a knowledge base, but also as a complex mix of "facts," texts, history, theoretical approaches, analytic frameworks, concepts, interpretations, and more. My goal, then, deals with "content" as it defines the study of literature. I further particularize this goal through class assignments and assessment when I assign specific texts for consideration.

2. IDENTIFY COMPONENT ABILITIES

Because the ability is a complex one, I need to break it open into component abilities. That step moves me toward the criteria to be used in judging it. In this case it means that my assessment of it does not occur in only one event. Instead, I consider how students can develop this ability throughout my course, and plan to assess for component skills in relation to specific texts students are reading.

One specific component skill of this overall goal is the ability to identify and discuss literary elements (such as plot, character, tone, style, and so on) with understanding of their use by different writers. This is a preliminary step to demonstrating how these elements work in relationship to each other as a reader develops a theory of meaning. I assess the preliminary steps so that I can assist students with difficulties they might have at this level before they attempt to demonstrate the more complex conceptual task.

The complexity or difficulty of the specific texts I assign also varies. As students develop the ability to analyze literary elements in stories that have more accessible styles or structures, I can assign more complex works to assess the extension of this analytic ability.

In effect, broader goals need to be broken open and spread on a continuum of development. For the beginning student, I set more specific skills to be developed. In an advanced literature course I would not assess students specifically for their ability to use the vocabulary of the discipline, although I presume this ability and in fact use it as a criterion for assessing a broader goal.

To determine an expected outcome for an assessment, then, an instructor needs to state an ability in relation to the learning content, the course context, the developmental level of the student, and the chronology of the assessment event.

3. SELECT OR DESIGN A STIMULUS AND CONTEXT

Although there are those educators who still bristle at the word "stimulus," it usefully describes that element of an assessment that elicits a student's performance. A stimulus might be a question asked, or an artifact presented for analysis, or a problem posed, or an event experienced. It might be a simple request for a choice of answers, or it might be a complex situation in which the possibilities for response are numerous.

Whether I choose a stimulus first and create a context for it, or begin with a context and then find an appropriate stimulus, I ask several questions. *How will I narrow the content to a concrete situation?* Will I be assigning specific texts, or events, or problems? A process or product or both? Will I ask the student to choose the specific content? How will I limit the choices the student has to make? *What do I want students to do with the content to show the ability?* Write paragraphs? Draw diagrams? Outline answers? Solve problems? *By what circumstances will I define limitations?* To what audience? For what purpose? With texts available to them? With time constraints? Alone or with others? *What will prompt them to do it?* How will I motivate my students to demonstrate all that they've mastered?

The Importance of Context. I want my students to be motivated to perform well on my assessment, to see their learning as part of their development as competent individuals. Establishing a realistic context is one way to do this, for it helps to break down some of the artificial barriers between the world of the classroom and other worlds. Most of us, as test designers, provide a stimulus by asking questions, but we rarely provide a context beyond the actual test conditions—open book, or timed, and so on. We ask students to write an essay, or to choose correct answers, or to solve a problem, or to remember data, without a setting or purpose that would relate their actions to anything else they might do.

A teacher might ask students to compare two authors, or presidents, or chemical compounds, without clarifying details that suggest why such a comparison is worth making. Students may or may not be able to imagine those details. And why should they be able to? Expecting them to do so often distracts them from essentials. If I ask students to make a comparison between two authors, it is my responsibility to provide a context. I do so when I ask them to make the comparison based on specific content and theory, for a specific audience, and for a reason: for example, to show inexperienced readers how language affects our response or to persuade a literary critic that his or her theory of fiction can be questioned with evidence from the authors being compared. It is also my responsibility to find ways of eliciting from them the elements of context they bring to their performance. In addition to examining their product, I must be able to perceive their assumptions and their reasons for particular emphases, selection of evidence, lines of argument, and conclusions.

Consideration of Developmental Levels. For the assessment designer, consideration of the developmental levels of the students plays as important a role in shaping stimulus and content as it does in articulating abilities to be assessed.

If I am teaching psychology, for example, I might assess beginning students on their ability to demonstrate understanding of several significant theories and advanced students on their ability to apply theories appropriately in actual situations. For the beginning student, a good assessment stimulus with context could be: "Create a detailed outline as a study-guide for other members of the class. What will they need to know to master the theories?" The difference here from a question that asks "Explain the major components of these three psychological theories" might seem small, but it is significant. Though the context is stimulated, it offers students a realistic purpose and audience to assist in selecting components of the theories and shaping an answer. It provides a framework that focuses the comparison and leads students to the task at hand. It relieves some of the time-absorbing activity of blindly determining a context for themselves—which is irrelevant to this situation and often misdirected. It relieves others of writing contextless prose consisting of nothing but generalizations addressed to no one. At its best, it enables them to see their learning as part of their development as competent individuals, able—and motivated—to perform.

From the advanced student, I expect a behavior beyond understanding psychological theories, so I provide a different kind of stimulus and context. Perhaps I will ask each student in a group to take on the role of a particular theorist and then pose specific situations to discuss from the unique position of that theorist and in dialogue with the others. Not only does this stimulus provide the opportunity for an interesting discussion, it also calls forth synthesis and application that carry understanding of the theories far beyond basic knowledge. The discussion mode lessens the students' control over the direction of thought, so that they can show whether they have sufficient understanding of the theories to apply them to whatever situation arises.

Consideration of Mode. In designing a stimulus and context, I maintain a major focus on developing a situation that will offer the student the best possible chance to show the particular ability for which I am assessing. If I am concerned with students' abilities to identify relationships between literary elements, for example, or to design a nutritional plan for a specific kind of patient, I must devise a "mode" that attempts to isolate these abilities. Rather than have students write essays, I might ask the literature students to construct a diagram or map of relationships within a particular story, and the nursing students might be asked to chart a plan. In such cases, students can demonstrate their ability to analyze relationships in a work of fiction or a nutritional plan without letting implicit demands for demonstrating other abilities, like clear writing, distract them.

Choosing a stimulus involves creating a leading question or situation; providing a setting and format for student response in ways appropriate to the specific outcome desired. It also involves recognizing that because a stimulus both shapes and is shaped by the integration of ability and content, my decisions about these elements and their connections can improve by becoming conscious ones.

4. DEVELOP CRITERIA

Whether one sees criteria as the standards by which one judges student performance or as the indicators of reasoning, judgment, values, and purposes by which one fills in their picture of a given ability, the process of developing them means inferring them from performances as experienced and remembered. In designing an assessment, a teacher will have in mind an "ideal performance." Though perhaps not consciously spelled out, it is part of the inspiration for the *need* for assessment.

My job as assessment designer, then, is to determine criteria by describing that ideal performance, distinguishing essentials, and generalizing enough to accommodate varied styles and varied qualities of performance. What was good about an analysis of a poem as it was done by a literary scholar, or by a student, or by myself? What was successful in a well-presented speech synthesizing several sociological theories? What made a review of a play valuable to a reader? I might consider performances I remember, or I might imagine a successful performance—another kind of "remembering." What would I want to see in a good analysis or a good review, or a good speech? My imagined idea of a good performance will probably be based on examples I have in my memory.

I can also determine or refine criteria by literally "collecting" performances. I build up a sense of what I can expect my students to accomplish as I see what students have actually accomplished.

Whether I base my criteria on remembered, imagined, or collected performances, I must be specific enough in defining criteria to allow the student (and myself as teacher) to recognize the ability that is being assessed.

In making them public, I not only assist the student, but also initiate a review process with my colleagues. Thus, these criteria become acceptable to other professionals in my field.

Specific Considerations in Identifying Criteria. Adapting criteria of performance to a specific instrument or process involves screening them with considerations of the content specified, the level of quality that defines the ability in a given context, and the developmental level of the student. If I want to assess students' understanding of particular aesthetic conceptual frameworks, I specify, "clarifies understanding of the relationship between art as an aesthetic construct and art as a reflection of life." If students have already sufficiently demonstrated understanding of specific frameworks, I might give them a chance to incorporate these frameworks into their own aesthetic perspective, for example, "clarifies understanding of relationships between selected aesthetic frameworks." If historical background is important to the ability I am assessing, I would add, "states the significance of relationships in terms of trends in literary history."

Providing criteria for levels of quality in a performance is perhaps the most difficult task for any instructor. So much of what we evaluate as "good" in a student's performance depends on unquantifiable feeling for the discipline. Just as students ask what makes one Shakespearian sonnet better than another, an instructor might well ask what makes one student's essay better than another, when both are adequate in the sense that they meet basic writing criteria. Perhaps both students have "clarified their assumptions about the artist's role," and both have "given evidence of mastery of the vocabulary of the arts." But one student is able to integrate the statement of assumptions into a coherent sense of the place of art in our society while the other student only lists assumptions. One student includes art terminology to highlight insights into particular works, while the other only uses vocabulary without error of definition. As I provide criteria for my students, I aim to detail statements of quality as well as more prescriptive criteria that define a basically acceptable performance.

The number of relationships might be specified for the less experienced student ("at least five") or made part of what is to be assessed for the experienced student ("relationships among major literary elements, major aesthetic frameworks, and major trends in literary history"). In addition to the developmental level of the student, the entire assessment process of the course guides my decision. And a simple consideration like time allotted for the assessment can keep my decisions realistic in relation to context as well as ideal performance.

Carefully spelled-out criteria for each assessment are necessary. Whether or not I *articulate* criteria, I continue to use them and to rely on my expert judgment when I assess student performance. We are convinced that taking the student seriously as learner involves making the basis for this expert judgment both available and refineable by articulating it in the form of criteria for each assessment. How fine or full a picture of expected outcomes individual teachers draw remains a function of individual experience and commitment. The traditionally difficult task of designing good tests and "correcting" them cannot be made easier by an assessment framework, but the difficulty can be rewarded by increasingly visible student learning.

5. PROVIDE FOR SELF-ASSESSMENT

If I aim to help my students take responsibility for their own development, I include a dimension beyond their demonstration of a given ability: I ask them to evaluate that demonstration. By designing criteria, I have provided them with the most important tools for self-assessment. But I still need to provide a time and a stimulus/format for self-assessment. I might include an overall question about the performance or a set of detailed questions about specific aspects of the performance. I might make self-assessment a formal part of the instrument or provide for it more informally through directive suggestions or questions.

Again, the key determining factors for my decisions are the level of the student and the context of the specific assessment. Where is the student in the development of his or her ability to self-assess? How does she use criteria? Does she have an internalized set? Does she have at least the start of a picture of her own strengths and weaknesses in regard to what is being assessed?

Before considering the range of formats for self-assessment mentioned earlier in this paper, I might decide whether to pitch the self-assessment to a more affective or a more cognitive level. Beginning students, in particular, might benefit most from a question that asks them to identify those aspects of the assessment they were able to handle with assurance and those of which they were unsure. At other times, they might best learn from a request to describe where they had a breakthrough in their thinking while they were working on the assessment. For more advanced students, I might ask for self-assessment in a more open-ended way, allowing them to supply their own categories.

In self-assessment, effective use of criteria remains a useful way for students to see an ability as a whole and to work on it in parts. Sometimes students can best use criteria as checklists to provide a profile of their ability. Sometimes they can best use them individually as take-off points for further understanding of their ability. In either case, I try to see that my individual assessments make provisions for self-assessment within the student's developing picture of her actual ability in relation to her potential.

6. JUDGE THE PERFORMANCE AND GIVE FEEDBACK

Judging performance and giving structured feedback constitute major elements of the assessment design. For the student, these may indeed be the most significant, since judgment and feedback are the visible signs of student progress, or the lack of it.

In the assessment design process we have been describing, judgment of the performance is a direct application of developing explicit criteria. As an assessor, I make observations of my student's performance and either record examples of the behavior I observe or at least mentally acknowledge them. On the basis of such evidence, I then judge the student's performance as it meets the criteria I have established. In the context of a course, I would also relate the student's performance to overall development of my course goals. When designing an assessment, I should think ahead to how, within the limitations of my time, I can provide feedback that will most benefit the students. Most importantly, I need to generate alternatives: written feedback in a checklist with one focused comment? written as a memo? oral on a tape recorder? oral in face-to-face interviews that replace several lecture periods? in combination with peer feedback? Whatever mode I choose, I generate feedback that provides students with a description of how they have performed. I describe for the student the successes that I find in the performance even as I make suggestions for ways the performance can be improved. A single criterion measures the quality of any feedback I give: Does it add to the student's dynamic picture of his or her own ability in a way that motivates further development?

Designing External General Assessments: An Extended Example

Extending assessment beyond the individual classroom to a wider curriculum context involves collaborative, integrating work by a group of designers from a single discipline or from several departments. Except for those additional factors, however, the design process is the same as for the classroom: determining expected outcomes, breaking each outcome into component abilities, creating an instrument, and identifying criteria of performance.

At Alverno, faculty have found it productive to design comprehensive assessments collaboratively. These instruments give them an opportunity to keep clarifying what they mean by general education or by specialization in particular majors. One such assessment is taken by Alverno students near the end of their second year. It is designed to give them a picture of how, on a given day in a given situation, they are able to bring together the abilities they have developed thus far. Faculty see it as another way of looking at each student's achievement in general education. By describing this assessment, we can illustrate the challenges and successes of a collaborative effort at assessment design.[7]

The assessment was originally designed by a general education committee. The group had set itself the task of selecting an instrument from which faculty could learn something about each student, and from which each student could learn something that would assist her in planning upper-division work in her major. They could not find an instrument that directly addressed some of the learning goals they had identified—aesthetic response, for example, or integrating observations and inferences to clarify meaning in a work or process. Therefore they designed their own, gradually working out a validity study with the Alverno Office of Research and Evaluation.

Abilities. The design process began with the abilities to be demonstrated. The committee took an ability like *aesthetic response*, for instance, and broke it into workable components, like making judgments about the quality of artistic works and defending judgments on the basis of how an artist sustains audience participation. They agreed on these as important aspects of student performance. After analyzing other expected abilities in a similar manner, they considered whether students would be able to integrate their abilities in working on problems that are not nearly separated into steps and do not come as the direct culmination of preparatory learning experience. Therefore the designers decided on a simulation, placing students in the role of a citizen advisory council to a local school board on the question of censorship of books. They imagined an entire set of tasks involving interaction with parents, teachers, and news reporters, as well as reading material on academic freedom.

Stimulus and Context. Gradually, a half-day assessment took shape that required each student to read background materials, and to prepare and deliver an oral presentation; to deal with a desk full of letters, phone messages, and memos by delegating or providing responses; and finally to develop, in collaboration with four other students, a set of guidelines and recommendations for which each must present a rationale. As soon as the faculty designers had the scheme sufficiently completed, they could assign the task of writing imagined scenarios to a creative, articulate teaching assistant and save for themselves the crucial task of identifying criteria of performance.

Examples of Criteria. Because the assessment was to focus on the integration of general education outcomes (ability and content), the faculty designers decided to aim for integrated statements of performance criteria. Two examples show the results:

1. Clearly articulates own position on issue (integrates valuing in decision making, communication)
2. Identifies implications of and rationale for own position, with accurate reference to and interpretation of a conceptual framework of one of the disciplines studied (integrates content, analytic ability, valuing in decision making)

Assessing and Administering. Since the assessment integrated and transcended course outcomes, faculty decided that student performances would be judged by external teams—volunteer professionals from the urban community, teaching assistants, and/or rotating faculty. The assessment would be administered during final assessment week in a situation external to any course. The assessors would also provide written and oral feedback, and self-assessment would be part of the feedback session.

This collaboratively designed assessment has been used successfully for more than ten years at Alverno. The results of the assessment provide ongoing diagnostic and summative feedback to general education instructors and to major departments, as well as to individual learners.

Assessing a Major in a Discipline. In any institution of higher education—whether or not it makes a total commitment to assessment—individual disciplines and departments have their own kind of opportunity. They can design unique assessments that address the abilities a student majoring in their field is expected to develop. An English faculty can have students act as members of a simulated civic cultural center for one week and assess their ability to evaluate literary materials from varied frameworks, work as members of an editorial board, and participate in an interview on significant literary trends. Like a ten-question comprehensive examination, the simulated assessment elicits students' knowledge and understandings. It also assesses their ability to reinterpret their knowledge and understanding in interactive situations similar to those they will experience as professionals. Behavioral science departments can create a simulated consulting firm or a research or clinical center. Traditionally, music recitals and art exhibits have provided culminating evidence, and celebration, of developed abilities. Other departments can learn from art and music how to build such dimensions into their assessments; at the same time, fine arts areas can extend assessments of recitals and exhibits by identifying explicit outcomes and criteria, by adding tasks that elicit additional abilities, and by providing vehicles for feedback and meaningful self-assessment.

Conclusion

We have emphasized the individual teacher as assessment designer and judge of performance, and have emphasized the potential of individual departments. Teachers or departments singly can try any of the strategies we suggest in order to experience advances in student learning. They can set course goals with a clearer focus on the learner, for instance, and organize instruction and assessment around the goals. Or they can be more explicit with students about learning goals and standards by which student performance will be judged. Teachers or departments can adapt other single aspects of the assessment-as-learning process we have described. They can provide learners with class-time practice in the use of goal-related abilities. They can improve their exams and their feedback by relating them more explicitly to learning goals. We believe that any of these strategies can of themselves make an immeasurably helpful inroad for a learner into the unmapped territory of his or her development.

However, it is not enough for individual faculty or departments to act alone. To work for the learner, assessment calls for a strong series of connections: expected outcomes must connect to criteria for performance, to assessment processes, to instructional strategies. On a day-to-day basis, these connections translate into relatedness between what students learn, how they learn, how they will be judged, and what their learning means for their future. In a collegiate institution, we consider the extent of those connections an important measure of the extent to which the environment is organized for learning. We might make some of these connections in a single course or program. But the more assessment is at the heart of the institution itself, the more its power can serve the learner.

Notes

1. Although the authors named are immediately responsible for this paper, we are indebted to the cumulative thinking of all of our colleagues at Alverno, especially other members of the Assessment Council: Zita Allen, Kathleen Bultman, Margaret Earley, Joyce Fey, George Gurria, Patricia Jensen, Wendell Kringen, Marcia Mentkowski, Glen Rogers, Judeen Schulte, Judith Stanley, Marilyn Thanos, Christine Trimberger, and Allen Wutzdorff.

2. The current level of understanding in regard to the question of teaching and learning for college students calls for careful observation, recording, and analysis of what is happening in specific contexts. Out of such studies will come questions for the synthesizing and experimenting stages of research. We see this paper as a contribution to the descriptive, analytic stage. Our propositions are based on our cumulative experience with students. In collaboration with our college's faculty as a whole, we continue to test our theory in the classroom and through ongoing institutional research.

3. Particularly helpful articles or chapters on shifting trends in testing and assessment are Robert Glaser, "A Research Agenda for Cognitive Psychology and Psychometrics," *American Psychologist* 36 (September 1981): 923–936; David C. McClelland, "Testing for Competence Rather Than for 'Intelligence'," *American Psychologist* 28 (January 1973): 1–14; Warren W. Willingham, "New Methods and Directions in Achievement Measurement," *New Directions for Testing and Measurement: Measuring Achievement: Progress Over a Decade*, no. 5 (San Francisco: Jossey-Bass, 1980).

4. There are numerous publications on the assessment center method in business. For a helpful overview see George C. Thornton III and William C. Byham, *Assessment Centers and Managerial Performance* (New York: Academic Press, 1982); and Joseph L. Moses and William C. Byham, eds., *Applying the Assessment Center Method* (New York: Pergamon Press, 1977). The former has an extensive bibliography.

5. For a detailed picture of assessment in Great Britain, see John Heywood, *Assessment in Higher Education* (London: John Wiley & Sons, 1977). Heywood's work includes a comprehensive bibliography.

6. For further examples of student responses, see the above publications and M. Mentkowski and A. Doherty, *Careering After College: Establishing the Validity of Abilities Learned in College for Later Careering and Professional Performance. Final Report to the National Institute of Education: Overview and Summary* (Milwaukee: Alverno Productions, 1984, c1983). A complete list of publications is available from: The Alverno Institute, Alverno College, 3401 South 39 Street, Milwaukee, WI 53215.

7. For further concrete examples of actual assessments, see *Assessment at Alverno College* by Alverno College Faculty (Milwaukee: Alverno Productions, 1985, revised edition). Other sources are G. Loacker, L. Cromwell, J. Fey, and D. Rutherford, *Analysis and Communication at Alverno: An Approach to Critical Thinking* (Milwaukee: Alverno Productions, 1984) and M. Earley, M. Mentkowski, and J. Schafer, *Valuing at Alverno: The Valuing Process in Liberal Education* (Milwaukee: Alverno Productions, 1980).

Assessment in Career-Oriented Education

by Sandra E. Elman and Ernest A. Lynton

The basic purpose of this paper is to draw attention to the nature and the role of assessment in career-oriented education at the undergraduate and graduate levels. There are two reasons this consideration is important. First is the sheer numbers and growth in this sector of higher education. In recent years, two-thirds of all baccalaureate degrees have been awarded in career-oriented curricula; the proportion at the master's level is even higher. Futhermore, the most recent survey of the Cooperative Institutional Research Program indicates that 27 percent of 1985 freshmen planned to major in business, compared to 19 percent in 1975.

Any general discussion of assessment of student progress and achievement in colleges and universities must, therefore, include areas such as business and management, engineering, nursing and many other health-related areas, teacher education, law, and medicine. Moreover, many of the pertinent issues need to be considered in the growing system of in-service instruction aimed at maintaining the competence of practitioners in the face of continuous change.

There is a second, very important reason to urge a critical look at assessment in career-oriented instruction. Assessment, in essence, provides a measure of how effectively someone has learned what has been taught. At this time, serious doubts are being voiced about whether what is being taught is really what students should learn. The criticism goes well beyond curricular details and raises questions about the basic approach to career preparation. We are facing an interesting chicken-and-egg situation: on the one hand, the mode and emphasis of assessment reflect what is being taught and therefore should change as a consequence of educational adaptations. On the other hand, assessment often provides a target for what is being taught. Perhaps changes in the assessment of career-oriented education—and perhaps also in the assessment of eligibility to practice—can be used to bring about the necessary modifications in the approach and content of career-oriented education.

This paper will describe the questions being raised about career-oriented education, and suggest the changes in assessment that would follow from and hasten curricular adaptations.

Current Criticism of Career-Oriented Education

One hears many complaints these days about allegedly excessive vocationalism in higher education. Yet it would seem that in abandoning the aims of a liberal education, our colleges and universities have also failed to be successful in preparing their students to be effective in a future occupation. Undergraduate and graduate programs aimed at preparing for a career are also being criticized. A few themes dominate: The curriculum is too narrowly confined to technical skills, there is too much of a gap between theory and practice, there is too much emphasis on purely cognitive and analytical material, and there is too much abstract classroom work and too little hands-on experience.

Most of these comments echo what Jencks and Riesman wrote twenty years ago. They pointed out the low correlation between course grades and occupational success (1969, p. 205) and described at length how the affiliation of professional schools with universities has, over the years, tended to deemphasize the school's occupational commitments and encouraged "a more academic and less practical view of what . . . students need to know." (op. cit., p. 252). They spoke of "the divergence between professional training and professional practice" and suggested that, just as undergraduate liberal arts units during the post-war years became "university colleges" with curricula directed toward graduate work in the disciplines, so also have professional schools focused more on "turning out men with skills appropriate to teachers [of the profession]," simply taking for granted that "these skills will also be appropriate to the practice [of the profession]." (op. cit.,

p. 253). Changing the name of several engineering and business schools to colleges of "engineering science" and "management science" was a striking symptom of this strong trend toward a more academic and abstract cast of career-oriented curricula.

Schein (1972) similarly commented on the narrowing, and, indeed, fragmentation of professional curricula. He stated that the professions have become so specialized as to become

> . . . unresponsive to certain classes of social problems that require an interdisciplinary and interprofessional point of view.
>
> Professional education provides no training for those graduates who wish to work as members of and become managers of intra- and interprofessional project teams working on complex social problems.
>
> Professional education generally underutilizes the applied behavioral sciences, especially in helping professionals to increase their self-insight, their ability to diagnose and manage client relationships and complex social problems, their ability to sort out the ethical and value issues inherent in their professional role, and their ability to continue to learn throughout their career. (Schein, 1972, p. 60)

Criticisms about the divergence between professional preparation and professional practice; narrow specialization; excessive emphasis on technical skills and cognitive factors; and lack of breadth all are once again being heard. Indeed, the need for a more practice-oriented approach, with less emphasis on the accumulation of facts, has become greater than ever. Practitioners must be prepared to deal with the new and more difficult job requirements created in most occupations because of rapid change and the complexity and interconnectedness of modern society. The ability to tolerate and to deal with ambiguity, to cope with disequilibria and discontinuity, to balance conflicting values and to assess risks, as well as to take risks all have become important conditions of functioning effectively in the contemporary context. The real world is messy, and there are few situations and problems that lend themselves either to clear definitions or to straightforward and unequivocal solutions.

Two Examples

Engineering provides one good example of the unprecedented challenges posed by the complexities of modern society and its technological advances. Competent engineers must have much more than scientific and technical skills. Increasingly, they should be familiar with the way in which science and technology operate in society. They need to realize that the ramifications and implications of their decisions have far-reaching consequences, many of which may be uncertain or even unpredictable.

As technical experts, they may be able to forecast with some degree of accuracy the first-order implications of a particular course of action, but that is not enough. Their analysis must also take into account the second- and third-order consequences that may have a direct impact on individuals, the environment, or perhaps the political and economic structure. To do so is very difficult. In addition to being only partially predictable, the second-order effects usually indicate the need to choose among competing values and objectives. That is what Prewitt (1983) has called the "bittersweet" principle of technological change. Technological innovation not only offers new social benefits; it also imposes social costs. Even small projects often undermine some social value, harm some social interest, and penalize some groups. At a minimum, most new construction requires some dislocation; most new techniques take away some jobs. Engineers must be trained to think about these matters and to develop a mind set that allows for a fusion of technical and other considerations, including ethical concerns. It is not enough for engineering students to master *technical skills*; they need to develop *technical judgment* (Jerath, 1983).

For managers as well, competence requires considerably more than mastery of technical skills. For one thing, it is increasingly important that even lower-level supervisors and managers acquire a better understanding of the context in which they function. Like engineers, they should learn to assess the second- and third-order effects of their decisions. The need for this skill is growing throughout the managerial hierarchy because of the current trend toward a more decentralized organizational style in which there is more delegation of authority and more shared decision making. A survey by the prestigious Conference Board (Lusterman, 1981) reported widespread agreement among corporate leadership that managers at all levels require competences such as:

- An awareness that events in the business environment significantly affect company interests and alertness to particular threats and opportunities;
- Sensitivity to how company decisions will affect, and be perceived, by others;
- Attentiveness to the opinions, values, and interests of others;
- An ability to systematically monitor and analyze the business environment, and integrate the data developed, into strategic planning processes." (op. cit. p. 6)

A further dimension of managerial competence derives from the changes in management style recommended by authors such as Peters and Waterman (1982), Reich (1983), Hayes and Abernathy (1980) and most recently Piore and Sabel (1984), who blame much of the decline of this country's international competitiveness in some fields on an adherence to the traditional, rigid principles of "scientific management." The suggested modifications and remedies differ in vocabulary and to some extent in substance. But all of these authors call for a management style that is more intuitive and more flexible, that tolerates ambiguity and accepts "messiness."

A New Concept of Practice

The new demands on the practitioner require basic changes in career-oriented education that go beyond a mere reshuffling of the curriculum. Broadening the program by the inclusion of a larger number of pertinent liberal arts subjects and by adding problem-centered, multi-disciplinary courses will be necessary. But this strategy is not sufficient to help students develop the kind of judgment required for good practice and to acquire the ability to deal with complexity and ambiguity. That calls for a rethinking and revision of the basic approach to career-oriented education.

In spite of wide use of clinical and other practical components, the pervasive emphasis in professional curricula continues to be on *content* rather than on *process*, on the *acquisition* of a body of knowledge rather than on the ability to *use* it. The current educational approach reflects the traditional view of professional practice as the systematic application of a set of standardized concepts and analytical methods to a recurrent problem in order to arrive at a unique solution. This positivist definition has become the hallmark of a profession. During the past decades, more and more occupations have been striving to achieve professional status by accepting this approach, which sets up a hierarchy of knowledge and a corresponding hierarchy of activity. Schein (1972, p. 43) has described the three components of professional knowledge:

1. an underlying basic science or discipline component which provides the fundamental principles of the practice;
2. an applied science or engineering component, which furnishes many of the diagnostic and problem-solving procedures; and
3. a skills component which consists of acquiring the ability to use the basic and applied knowledge in actual practice.

The application of 1 yields 2, and in turn that leads to 3. As Schön has pointed out,

> . . . the order of application is also an order of derivation and dependence. Applied science is said to 'rest on' the foundation of basic science. And the more basic and general the knowledge, the higher the status of its producer. (1983, p. 24)

This hierarchy is reflected in the basic structure of current career-oriented programs. Even in fields that can lay only a marginal claim on professional status, the curriculum usually begins with what are viewed as the pertinent basic sciences. These are followed by a number of applied science and technology courses. The curriculum ends with clinical experiences intended to provide opportunities to develop skills of application (cf. Schein, op. cit., p. 44). Throughout, *learning precedes doing*, and *practice is viewed as the application of theory*. That is the model which, particularly since World War II, has become normative for almost all career education.

There are good reasons to believe that this traditional positivist approach is no longer adequate. When Ackoff (1979) speaks of "managing messes," he describes situations for which no technique provides a single and direct path to a unique solution. In most cases there are likely to be several alternatives, each with its combination of advantages and disadvantages. Exercising "technical judgment" or "managerial judgment" in such situations is a rather different process from the traditional application of predetermined techniques. The majority of situations faced in daily practice of most occupations cannot be readily reduced to the application of standardized problem solving methods. Indeed, *problem definition and clarification*, rather than *problem solving*, emerge as the major tasks.

Schön (1983) believes that an effective practitioner approaches each problem

> . . . as a unique case. He does not act as though he has no relevant prior experience, on the contrary. But he attends to the peculiarities of the situation at hand. . . . [He does not behave] as though he were looking for cues to a standard solution. Rather [he] seeks to discover the particular features of his problematic situation, and from their gradual discovery, designs an intervention. (p. 129)

The title of Schön's book, *The Reflective Practitioner*, describes his basic view: successful practitioners learn while doing. They engage in what Schön calls "reflection-in-action" as they interact with their client or with the situation they are facing. It is, in essence, an ongoing feedback process of successive approximation, of which the architectural design process is an excellent example.

Implications for Career Education

This radically different view of professional activity suggests, as well, a substantial change in career education. The crucial need is to use simulated as well as real experience in very different ways from what is currently done. Instead of constituting "practice"—that is, merely ways of acquiring the skill to apply prior learning—the experiential components of the curriculum must become themselves primary learning devices. Learning must be related to and derived from doing, instead of preceding it. Greater emphasis is needed on inductive reasoning and the power to generalize. Both sequence and hierarchy of the curricular components of career education must change, with the clinical and other experiential parts occupying both a more pervasive, as well as a more important, place.

We point out elsewhere (Lynton and Elman, 1986) that the dichotomy between liberal and career-oriented education is false and dysfunctional because the two have substantially analogous objectives. Both should emphasize *competence*. Competence on the job and competence as a member of society both involve risk assessment and risk taking, striking balances between competing values, and a shift from answering questions to deciding which are the right questions to ask. But such a similarity of goals does not imply congruence of curriculum. It does not mean that an undergraduate major in an arts and science subject is the best preparation for an occupation. That view, so frequently expressed these days, totally denigrates the continuing need for occupation-specific expertise. Competence on the job requires *more* than technical skills, but it does include such skills. Process is vital, but it cannot be empty of content. Competence transcends knowledge, but must include it.

The Implications for Assessment

If simulated and real experiences are to become major sources of learning in career education, if a principal goal of such education is to enhance the competence of individuals in the practice of their occupation, if *process* is to become as important as *content*, then these emphases must be reflected in the assessment of student progress and achievement. At this time, the preponderance of assessment in career education—as well as in all other programs—is of the most traditional kind: written course and comprehensive examinations which test the students' grasp of basic principles and of pertinent facts. Such paper-and-pencil exams tend to be used even in clinical courses. The subject of negotiation is a typical example. The majority of business programs include it, usually with several opportunities for active student involvement in simulated negotiating sessions. In many institutions, these are video-taped and provide a useful source of self-assessment. But when it comes to assigning a grade to the student for the course, most instructors rely on testing knowledge of textbook material. Much the same situation exists in the clinical components of other career programs, such as patient interviewing and diagnosis in social work and medical education, or moot court activity in legal education.

Traditional assessment of factual knowledge and analytical skills continues to be important and must remain an important part of career education. But there also needs to be substantial assessment of experiential performance. Perhaps a better way of putting this is to use the basic distinction between testing and assessment. In their paper, Loacker et al. (1986) state this distinction very clearly:

> Testing can tell us how much and what kind of knowledge someone *has*. Assessment gives us a basis for inferring what that person can *do* with that knowledge.

Thus what is needed in career education—as in all other education—is a move from a virtually exclusive emphasis on *testing* to a substantial inclusion of *assessment*.

Current Trends in Licensing and Certification

The modes of assessment currently used for licensing and certification are almost entirely *content* oriented. They test the acquisition of pertinent knowledge rather than the ability to use and transmit that knowledge. The inadequacy of this approach is particularly pronounced in fields in which the tested content, i.e. the basic body of knowledge and methodology by which the profession defines itself, does not have a firm theoretical grounding and is subject to criticism. A striking example of this dilemma is the area of teacher education. Much of the current criticism of the conditions of our schools raises questions about the pedagogical theories and other bodies of knowledge taught in schools of education. In an effort to find some feasible solutions,—that is, produce better-trained teachers—many states currently are seeking alternatives to certification requirements based primarily on testing classroom knowledge of the traditional subjects in education.

The search for alternatives is proceeding in two very different directions. The first substitutes one kind of content for another, replacing knowledge of educational theory and methodology with knowledge of the specific subject matter to be taught. New Jersey, for example, now offers the first district-administered training programs leading to teacher certification. Those districts with such programs will have the authority to hire on provisional contracts college graduates who have passed competency tests in the subject areas they will teach but who have not been certified through traditional education programs. In addition, these districts are authorized to recommend those individuals who successfully complete the district-administered programs for state certification.

A very different direction—and in our opinion a more valid one—is the move toward basing teacher certification, in part, on demonstration of competence in the classroom. That such assessment is possible on a systematic and large scale basis is demonstrated by an innovative process developed by the AMA in the area of management, using principles and approaches that would seem applicable to classroom teaching and other occupations.

There also exist pervasive efforts to modify recertification policies in teacher education. A recent survey indicates that twenty-nine of the forty states that require recertification will allow teachers to meet some of the requirements by participating in in-service training sponsored by local school districts. Eighteen of those twenty-nine states now allow all of their recertification requirements to be met at the district level. The rise of district-planned recertification reflects a growing sense that traditional campus-based courses are not adequately meeting the staff development needs of individual school districts, and that the means of instructing teachers and assessing their skills and knowledge are not adequately focusing on actual professional practice. (Hanes and Rowls, 1984)

Performance-Based Assessment

The outlook for more emphasis on performance-based assessment in career-oriented programs is good. Almost all pre-professional curricula incorporate real or simulated clinical experiences: internships and clerkships, case studies, moot courts, and a variety of simulated games and role playing. The amount of this is increasing, and there exist more and more instances of practical experiences being incorporated into the early stages of a student's curriculum. However, in most programs these activities continue to take place during the final phase of a student's course of study. The maxim "theory should precede practice" is still paramount. And, for the most part, clinical periods are viewed as additional components to, rather than as integral parts of, the academic program. Clinical experiences are considered to be opportunities to practice prior learning, rather than sources of new understanding in and of themselves.

If performance-based assessment is to become an important element of career-oriented education, it is necessary to incorporate real or simulated experiences as earlier and more integral components of the curriculum, and to structure them in such a way as to provide valid opportunities for assessment. The following section will describe two basic elements that are necessary for this purpose.

Facilitating Assessment: Structure and Process

Small Group Interaction. For a real or simulated period of professional practice to be a primary source of learning as well as an opportunity for assessment, both faculty and students must focus on *process* as well as *outcome*. To be sure, it is important that the faculty member observe a student's performance. Yet it is important to go one step further. The faculty needs to *understand* the student's rationale for making certain decisions and behaving in particular ways. One way of doing this is to build into the moot court, case study, and internship experience the component of small group interaction.

Small group interactions provide opportunities for assessment particularly in such career-oriented curricula as law, teacher education, medicine, and nursing, and to a lesser but growing extent, in business administration and engineering. They are particularly useful whenever the problems encountered can be subject to different interpretations and alternative outcomes. Future professionals need to understand and evaluate their own capacity to make sound judgments and to display professional expertise. Such understanding and evaluation may be enhanced by providing students with a setting in which they can express and explain the rationale for their behavior or anticipated behavior.

By engaging in such a dialectic process, faculty and peers are able to probe a student to elicit the cognitive and non-cognitive chain of events that led up to a certain decision. By mapping out one's line of reasoning, both the student and the assessor can better measure performance. Small groups provide a conducive setting for eliciting ongoing explication of a student's thought processes that lead to certain decisions and behavior. The structure of the small group and the intensive and intimate information-exchange process allows the assessor to go beyond the surface in judging performance.

The assessors (the "expert" judges in the small group) would be the faculty member(s) as well as the student's peers. The criteria they would use to judge the competence of the student would be developed by the faculty (and perhaps the students) before the small group sessions. The criteria would vary from program to program and perhaps even within programs.

The small group experience, as a means of not only training but assessing students, is particularly useful in such high-pressured professional fields as medicine, where often the "problems of life and death are presented to the students without sufficient preparation and without giving them the opportunity to influence or examine their human responses to such basic experiences." (Neumann & Elizur, 1979, p. 714). The sole purpose of establishing small groups need not be as a means to assess performance. On the contrary, the objective should be two-fold: instruction and assessment. The two phenomena should be viewed as interdependent.

Expert Judgment. One of the most overlooked yet critical components of assessment is that of expert judgment. The role of the expert in assessing student performance has special significance in career-oriented education. An earlier section pointed out that effective professional performance constitutes a synthesis of technical skills through the application of technical judgment. It follows then that the assessment of performance in these fields must likewise be the embodiment of a synthesis of the judgment of those capabilities. The assessment process, therefore, cannot limit itself to the component skills, but must systematically examine the interplay of these related factors. As a result, the assessment of performance is not a purely objective or quantifiable exercise. By its very nature, it incorporates and reflects subjective analysis and normative values. Furthermore, an element of uncertainty is inevitable in the assessment process. As the scope of technical decisions becomes increasingly complex, the degree of precision in ascertaining their validity decreases. The expert judgment in assessing student performance must obviously be applied by individuals who have first-hand familiarity with professional practice.

We have argued elsewhere (Lynton and Elman, op.cit.) that it is essential for faculty in career-oriented programs to have such first-hand experience in order to be effective curriculum designers and instructors. The need for faculty members to exercise expert judgment in student assessment adds another degree of urgency to making them more familiar with actual practice, and enhances, as well, the usefulness of practitioners as adjunct faculty.

In addition, it is advisable to use external experts in the assessment of student performance. To do so does not eliminate ambiguity and subjectivity in the assessment, but it greatly enhances the validity and reliability of the evaluation. However, it is not always easy to identify expert judges among practitioners. The fact that an individual holds a certain professional title and occupies an office affiliated with a prestigious institution, firm, or clinic does not necessarily insure any particular level of expertise. To a large extent, expert judges are identified as such by their peers who regard their work as being of superior quality and having a significant impact in their field over a period of time. The criteria by which these individuals are judged to be experts undoubtedly varies from one professional field to another. Faculty who engage in applied research, technical assistance and policy analysis themselves, and who maintain linkages with fellow professionals beyond academe, are more likely to be in a position to tap such expert resources and bring them into the assessment process.

The expert judge knows that effective professional practice depends upon a continuous process of questioning and evaluating one's own actions in light of changing technologies and ethical and normative imperatives. That is why one of the most crucial components of career-oriented education and the assessment process is self-assessment.

Self-Assessment

''Know thyself'' may be a maxim primarily associated with philosophical inquiry, but its pertinence to professional practice should not be underestimated. Understanding and being able to evaluate one's own actions are essential to effective professional practice. There is no straight path from a well-specified problem to a unique solution. Rather, as Schön has pointed out so clearly, the effective professional engages in a continuous process of trial and error, with ongoing feedback that provides guidelines for improving the quality of one's actions.

If career-oriented education is to inculcate the lifelong habit of self-assessment in one's occupation, such introspection must be included as an integral part of the curriculum. The preceding section indicated that a valid process of performance assessment, using small group interaction and expert judgment, automatically contains a strong element of self-assessment.

What then does doing effective self-assessment mean? Self-assessment implies that an individual is engaged in several cognitive and affective activities that concern one's self, education, and professional preparation. These may include: defining goals; identifying personal strengths, weaknesses, skills, knowledge and intents in different roles; and acknowledging problems and seeking resources for help. (Withorn, 1982, p. 14)

Much of the value of self-assessment, if done effectively, is that it increases students' awareness of what they are learning, and more importantly, the relationship of that knowledge and skills to future tasks. When a student asks, ''What have I done, and how did I respond,'' he or she is creating both cognitive and affective relationships that ultimately make actions more reflective and less rote. Self-assessment encourages individuals to think about the normative ramifications of their decisions and to apply what they have learned from one experience to another.

Innovative Approaches

In essence the problem besetting educators in career-oriented education is how to assess performance that essentially is a synthesis of technical knowledge, technical skills, and technical judgment, and is inherently holistic. There may be no "right" measures for assessing performance in career-oriented education because, by their very nature, the activities are not performed as discrete units. Professional practice is not merely a series of acts, rather it is a process. It would follow that the assessment of professional performance must be process-oriented and holistic as well. Not all approaches to assessment, however, may reflect that notion. There may be no one best way to approach assessment in career-oriented education; in the elementary stages we may have to learn through a process of trial and error.

The American Management Association (AMA) and the Harvard Medical School have recently developed programs which pursue markedly different approaches to assessment of professional performance. The AMA assessment model seeks to determine the extent to which a student has acquired and can use eighteen generic management competencies which were designated common to superior managers by a research team that had reanalyzed over 2,000 job studies.

The AMA's competency model differs from other generic models in the criteria used in determining management competencies. Most other competency models are based on theories of management and findings of expert panels and/or job analyses. By contrast, the competencies in the AMA model were ascertained by analyzing the components of the performance of outstanding managers.

In this case, the criteria used for judging students' performance are not determined by the faculty, but rather by outside researchers; however, it is the faculty members who assess the student's level of competence. This assessment process has a dual focus: audit and feedback. The audit process involves: a) four interactive exercises with simulated recreations of varied managerial situations; and b) a battery of tests designed to evaluate students' traits, motives, learning styles, cognitive abilities, and interests. Knowledge competencies are tested by both objective and case study exams. In addition, video-taped exercises and an audio-taped interview are assessed by being analyzed and coded in terms of the basic competencies on which the program is based. The results are shared with the participants during the feedback process. The essential components of the audit and feedback activities include a competency profile based on input from tests, questionnaire results, and data regarding an individual's behavior patterns from peers and faculty. In addition, each student receives a "Development Plan," which is a blueprint for action to fill the gaps in knowledge and skills identified in the audit, as well as a "Back Home Simulation," which allows participants to apply what they have learned in simulated workplace situations.

Clearly, the AMA model embodies a strictly defined set of procedures that are quite rigidly adhered to in an effort to produce more competent managers. The AMA approach thus attempts to reduce the levels of ambiguity and uncertainty as much within the training process itself as the manager might seek to do in the workplace.

By contrast, the Oliver Wendell Holmes Society's New Pathway Project in General Medical Education at Harvard proceeds from the assumption that there is more uncertainty and ambiguity both in the process of training and assessing medical students and in the world of medicine than has been previously acknowledged. It is not the body of knowledge that is under scrutiny, but how to apply that knowledge. The AMA model implies that good management technique rests upon a framework of action that is rational and determinate. By contrast, the HMS model rests on the premise that good medical practice requires rational as well as intuitive (or what Schön would call artistic) judgment.

The New Pathway Project (which includes twenty-four randomly selected students in its first-class, 1985-86) is designed to address the critical needs and pressures of medical educators and students. Assessment is a central feature of the Program. Like many other aspects of the Program, the evaluation component is very much an *interactive* process. Faculty and students work closely together. A faculty advisory network closely monitors student progress and provides regular feedback to the student and preceptor. The preceptor, in turn, provides ongoing appraisal of the student's interpersonal, attitudinal, and skills development.

A conceptual framework, set forth in a list of "guiding questions" with accompanying references and support materials, directs the students to the key principles, concepts, and learning issues in each unit of the curriculum. Students are evaluated for their general knowledge, problem-solving, and clinical reasoning abilities by their responses to a selected set of these "guiding questions." The emphasis on content is in no way diminished by this approach. Mastery of essential knowledge is appraised by means of self-directed testing, and clinical competence is tested using programmed patients, allowing cross-student comparisons and assessment of a single student's development over time. Overall evaluation of students is competency-based: students respond to a randomly selected, statistically significant sample of the total set of "guiding questions."

The effectiveness of the New Pathway approach will itself be assessed by comparing the performance of students in the Program with those pursuing the standard curriculum on such factors as:

- knowledge of basic science and scientific method;
- clinical problem solving ability;
- modes of self-learning and self-assessment;
- professional attitudes; and
- adaptive strategies for coping with stress.

Students from both groups will be interviewed periodically to study the evolution of their concepts of competence and caring. (*Harvard Medical Alumni Bulletin*, 1984, pp. 14-24)

What makes the New Pathway's process of assessment so remarkable in terms of the history of medical student education, is the degree to which it reflects a new Gestalt in educating—and evaluating—the medical student. Traditionally, evaluation of a medical student's progress has been a formal, well-defined process aimed at measuring a predetermined set of outcomes primarily through written and oral examinations. There was little, if any, emphasis on attitudinal and behavioral factors, or on assessing performance and progress through interpersonal communications. Much of the "new wave" orientation within the New Pathway Project is not unique to Harvard. Similar innovations are taking place at the medical schools of McMaster, Brown, and Southern Illinois Universities.

A striking difference between the AMA's and Harvard's New Pathway's approaches to assessment is that the latter emphasizes a continuing, never ending process of *becoming* a professional practitioner, with the recognition that the decision outcomes and ultimate behavior may not follow any prescribed procedure, and that every demonstration of technical judgment may be unique because of the differences of each situation the professional encounters. The AMA's approach to assessment, by being based on eighteen well-defined generic competencies, implies a less open-ended, more determinate process, with a well-defined and replicable outcome.

In conclusion, we wish to reiterate our conviction that in both its content and its modes of student assessment, career-oriented education needs to place greater emphasis on performance under real or sited practice-related conditions. It is important to shift from an "information-intensive" approach to one that stresses the ability to use cognitive, as well as other forms of knowledge in complex and ambiguous situations. The similarities between effective practice in a broad range of professions far exceed the differences. What is valid and necessary for medical competence is largely applicable, as well, to professional practice in management, engineering, and many other fields. The kinds of innovations in assessment and self-assessment pioneered in a few medical schools and some management programs at this time should find their way, with appropriate modifications, into other career-oriented curricula as well. Given the proportion of our undergraduates and graduates enrolled in career-oriented programs, this issue should receive as much attention as that of assessment in the liberal arts.

References

Ackoff, Russell. "The Future of Operations Research Is Past," *Journal of Operations Research Society*, 30, 2, 1979, 93–104.

American Management Association, The Master of Management Program of the American Management Association. New York: American Management Association, no date.

Cooperman, Saul and Leo Klagholz. "New Jersey's Alternate Route to Certification," *Phi Delta Kappan*, 66, 10, June 1985, 691–695.

Feeney, E. "Performance Audit, Feedback and Positive Reinforcement," *Training and Development Journal*, 26, 1972. 8–13.

Feistritzer, C. Emily. *The Making of a Teacher: A Report on Teacher Education and Certification*. Washington, D.C.: National Center for Education Information, 1984.

Fisher, B. *Small Group Decision Making: Communication and the Group Process*. New York: McGraw Hill, 1974.

Gannon, Martin, J. *Organizational Behavior: A Managerial and Organizational Perspective*. Boston: Little, Brown & Company, 1979.

Hanes, Madlyn Levine and Michael D. Rowls, "Teacher Recertification: A Survey of the States," *Phi Delta Kappan*, 66, 2, October 1984, 123–126.

Hayes, Robert, H. & William J. Abernathy. "Managing One Way to Economic Decline," *Harvard Business Review*, July-August 1970, 67–77.

Harvard Medical Alumni Bulletin, "A New Pathway for General Medical Education," 58, 4, Winter 1984, 14–24.

Innovation in Learning: Realizing the Potential of American Higher Education, Washington, D.C., National Institute of Education, October 1984.

Jencks, Christopher and David Riesman. *The Academic Revolution*. Chicago: University of Chicago Press, 1968.

Jerath, Sukhuarsh. "Engineering Education in Perspective." *Mechanical Engineering*, 105, February 1983, 92–93.

Kuriloff, Arthur, H. *Organizational Development for Survival*. New York: American Management Association, 1972.

Loacker, Georgine; Lucy Cromwell, and Kathleen O'Brien. "Assessment in Higher Education: To Serve the Learner." Paper prepared for the Conference on Assessment in Higher Education at the University of South Carolina, Columbia, October 1985.

Ludmerer, Kenneth, M. *Learning to Heal: The Development of American Medical Education*. New York: Basic Books, 1985.

Lusterman, Seymour. "Managerial Competence: The Public Affairs Aspects," Report #805, New York: The Conference Board, 1981.

Lynton, Ernest, A. *The Missing Connection Between Business and the University*. New York: ACE/Macmillan, 1984.

Lynton, Ernest, A. & Sandra E. Elman. *The Extended University*. San Francisco: Jossey-Bass, Inc., forthcoming.

Mahon, Noreen, E. "Caring: An Ethical Issue" in Louis H. Orzack & Annell L. Simcoe. Eds. *The Professions and Ethics: Views and Realities in New Jersey*. Bureau of Educational Research and Development, Rutgers University, New Brunswick, New Jersey, 1983, 92–99.

Margolick, David. "The Trouble With America's Law Schools," *The New York Times*, May 23, 1983, Section 6, 20–25, 30, 32, 36, 38.

Moore, Wilbert. *The Professions*. New York: Russell Sage Foundation, 1970.

Neumann, Michael and Avner Elizur, "Group Experience as a Means of Training Medical Students," *Journal of Medical Education*, 54, 9, September 1979, 714–719.

New Pathway Project In General Medical Education. 1984-85 Summary Progress Report. Harvard Medical School, May 29, 1985.

Peters, Thomas, J. & Robert H. Waterman, Jr. *In Search of Excellence*. New York: Harper & Row, 1982.

Pfeiffer, J. William & John E. Jones, eds. *A Handbook of Structural Experiences for Human Relations Training*, Vol. I revised. La Jolla, CA: University Associates, 1974.

Physicians for the Twenty-First Century. The GPEP Report. Report of The Panel on the General Professional Education of the Physician and College Preparation for Medicine. Association of American Medical Colleges. Washington, D.C., 1984.

Piori, Michael J. and Charles F. Sabel. *The Second Industrial Divide*. New York: Basic Books, 1984.

Prewitt, Kenneth. "Scientific Illiteracy and Democratic Theory," *Daedalus*, Spring 1983.

Price, David. "How Can Links Between Ethics and Practice Improve?" in Orzack and Simcoe, eds. The Professions and Ethics, 88–91.

Purtilo, R. B. & C. K. Cassels. *Ethical Dimensions In The Health Professions*. Philadelphia: W. B. Saunders Company, 1981.

Reich, Robert, B. *The Next American Frontier*. New York Times Books, 1983.

Schein, Edgar, H. with Diane W. Kommers. *Professional Education*. New York: McGraw Hill, 1972.

Schön, Donald. *The Reflective Practitioner*. New York: Basic Books, 1983.

Shimberg, Benjamin. Occupational Licensing: A Public Perspective. Educational Testing Service Center for Occupational and Professional Assessment. Princeton, New Jersey, 1982.

TO IMAGINE AN ADVERB

Concluding Notes to Adversaries and Enthusiasts

by Clifford Adelman

This conclusion has three limited purposes: First, to speculate too briefly on the role of judgment in culture and language as a theoretical ground for thinking about assessment. Second, to indicate some key issues that are more implicit than explicit in the collection of essays you have just read. Lastly, to provide a bibliography of technical and theoretical references that institutions may find helpful in more advanced stages of their study of assessment.

Assessment: One Word, Two Worlds

As the papers in this collection have indicated, assessment is a comparatively new word in the language of higher education. It is a word with which we seem alternately fascinated and fearful. The emotive meaning with which we thus load the word is paradoxical, since higher education has been in the business of assessment since the founding of the medieval universities, and that business has been tacitly assumed by all cultures and economies. It is, by one interpretation, an indispensable system in the body of our institutions, much like our lymph nodes or bone marrow: we rarely think about it, but we couldn't live without it.

Having discovered the word, though, some have embraced it as the heart of higher education, wishing upon it functions it may not be designed to perform. Others have simply wished it away, believing it to be either extraneous or an outright threat to the body academic. While the essays in this volume both explain and explore the enthusiast's response, the first part of this conclusion will address the adversary's position.

Let us create two worlds by the two major uses of the term, "assessment." The first is that of ordinary language use, broad and informal: "assessment" is used almost interchangeably with "judgment," but the situations in which one would choose the term are those involving behavior, the results of behavior, or the possibilities for behavior.

The second use of the term is technical, codified, and formal: "assessment" is an umbrella term for the activities of gathering, measuring, and communicating information about individual human performance with respect to discrete tasks requiring the demonstration of knowledge and skills. These activities vary according to the intended uses and users of the information, for example, screening (e.g. college or program admissions), sorting (e.g. placement), enabling (e.g. as an instructional method), certifying (e.g. degree qualifying), etc., but always involve more than one occasion of measurement or more than one judge if they are to be assessments. Uses of aggregates of information based on individual performances are, in our common parlance, "evaluations."

In either world, one would have to alter human language rather radically in order to do away with assessment. Some languages have developed with a minimum of prepositions, but none to my knowledge has developed without adverbs. The adverb judges performance; it compares, places, and qualifies events and actions. It not only tells us where and when actions took place, but *how* they took place. Verbs do not stand in an undivided empire of meaning; it is an inherent tendency of the human mind to qualify observations of events, actions, performances, and in so doing, to judge.

Ancient epistemology, modern metaphysics, and contemporary psychology all reinforce this fundamental notion evident in the very nature of language. Judgments, as Kant pointed out, involve quantity, quality, and relations—*irrespective* of their particular empirical content. In our time, both Piagetian psychology and generative grammar draw on similar assumptions concerning innate forms of reasoning: there is something *a priori* in the human mind that enables it to order and judge experience, and that "something" can be induced logically or inferred from behavior, particularly verbal behavior.

The faculty of judgment, however, does not yield "assessment" until it is codified in social organization. This codification is inevitable, and arises from the needs of society for expertise. As the social unit becomes more complex, various kinds of "economic" activity emerge, each demanding specialization of behavior, and each specialized behavior involving an informal selection process based on multiple observation and judgment.

When it comes to matters that the social unit regards as critical to its survival, this informal judgment evolves into certification. The society formally confers a collective judgment of expertise through a public symbolic act. By this theory, the very first certification occured when we annointed our tribal priest. By whatever criteria we had established, this person was judged by us to be able to represent our interests and hopes before the gods. The act of annointing was our way of granting the priest a license, a public acknowledgement of expertise. And no doubt we authorized this person to carry artifacts that symbolized that license—not unlike the badges, uniforms, and parchments we grant to holders of licenses and similar certifications in a modern economy. In all these cases, though, ancient and modern, we grant the "license" by reference to criteria established by our culture and by comparing the qualifications of various individuals to hold that license.

In a simple truism, Eugene Webb (1966) reminds us that "measurement is . . . always . . . a comparison" (p. 6). It is a comparison of a representation of a reality to another representation or to the reality itself. Consider: when we use a bathroom scale, we compare the number indicated (a representation of reality) to our sense of a standard, and we judge whether the reality represented by the number should be more or less. The criteria for that standard are both psychological and empirical. Medical science says that at such-and-such a height, bone structure, age, and sex, ideal weights should range in a given band. How have these "ideals" been determined? By empirical studies of the relationship between these variables and indicators of health in large samples of the population. People with characteristics M, P, and Q who want to live longer or stay healthy, will strive to maintain a weight within the band. Of course, some people may wish to defy the odds. But the point is that they have norms, indicators for assessing where they should be, and they know the consequences of not being there.

There is not much difference between these norm-referenced measures and those applied in education. Through the accumulation of evidence and practice, scales and norms have been established in reading, for example, in virtually every culture in the world. Equally in India, Brazil, France, and the U.S., there are indicators of what it means to read at X level, and research has well demonstrated that what Cummins (1980) called cognitive-academic language proficiency (CALP) is one of the strongest determinants of an individual's academic development. Given the evidence, all of these systems say that a student at a particular level of education or in a particular institutional environment should be able to accomplish certain reading tasks in order to succeed at that level of education and in that environment. As in the case of body weight, we know the consequences of not achieving that particular level of reading.

These examples bring us to the second world of assessment, one that evolves naturally from the first. Codifications of judgment as to who does what *how* are inevitable when the economic order involves an increasing division of labor. We now live in a culture dominated by licensure and credentials, both of which require third-party assessment of our knowledge and performance. Not only does this system seem to work fairly well, but we all expect it to work. Our expectations are reflected in a web of laws, regulations, and guidelines we have implicitly demanded of governments and professions, and in the fact that most of the assessments lying behind credentials and licenses are absolute. That is, you either pass the assessment or you don't; and there is a rather unambiguous line of demarcation. We do not award half pilot's licenses; and we do not award full licenses to the person who passed flight training but not the navigation exam. The same type of observation can be offered for nurses, accountants, pharmacists, stockbrokers, architects, real estate agents, etc. Over 800 occupations are licensed in one or more states; and to be licensed in 500 of those occupations—from air conditioning mechanic to medical records technologist—one must pass a written examination (Wigdor and Garner, 1982, p. 133).

The issue goes beyond licensure. For example, Wigdor and Garner report that approximately 1.6 million applications are received every year for positions with the Federal government ranging from stenotypist to air traffic controller for which some kind of assessment is required and that approximately 45% of those assessments involve written tests (pp. 124-5). The process of selecting a foreign service officer, for example, involves written tests, a writing sample, an in-basket test, an interview, and a presentation/negotiation exercise. In the course of these assessments, some 27 categories of knowledge, skills, aptitudes, and personal characteristics are judged against pre-set criteria.

On the state level, the data cited by Wigdor and Garner show that at least 33 state governments use written tests to screen applicants for technical jobs, at least 30 do so for professional jobs, and at least 18 for managerial and administrative jobs (p. 129).

The point is that assessment is no passing fad in either society or education. An education system that neither predisposes nor prepares students to take "third-party" examinations, an education system that does not assist students in understanding and articulating criteria for performances based on cognitive skills, is condemning them to a life outside the economic mainstream of virtually every nation on earth, indeed, is condemning economies themselves to uncertainty and mediocrity.

In light of these realities, there is a profound paradox in American higher education. As the papers in this volume by Terry Hartle and John Harris point out, we are virtually the only major higher education system in the world that has combined the instructional and certifying function in the same person—the individual faculty member. Our custom of continuous classroom assessment by different faculty is—by the standards of most nations—bizarre, if not outrightly inefficient, since shared criteria for what makes for academic achievement are few. As the ACE Task Force on Credit and Credentials observed, classroom "assessment techniques vary from the crude and simple to the refined and sophisticated," and standards of performance "fluctuate according to faculty members and examination systems, the qualifications of students available, [and] the state of development" of a given field (Miller and Mills, 1978, p. 20). What is even more paradoxical—and tragic—is that it is not considered polite to talk about the differential criteria for achievement in American academic circles. The reason we have *de facto* national standardized tests at the point of entrance to graduate or professional school is that at least they represent a common currency—something we cannot say about college credits, grades, or degrees.

Credits and Credentials: The Faith of the Academy

Some of the impetus for the current assessment movement emerges from the broader credentialing function of institutions of higher education. The degree, as a credential, holds an "advisory" status, but some degrees are more than that, e.g. "one requirement in qualifying for governmental or voluntary credentials" (Miller and Mills, 1978, p. 9). We often overlook the fact that a great deal of assessment goes on in American higher education precisely because of those degrees that are requirements in the credentialing process in fields related to public health, safety, and welfare and/or in occupations for which either the state or a voluntary professional association requires a license or certificate.

The majority of degrees we award, however, do not fall in these fields. The credentials may be in the public interest because they "recognize and encourage pride in accomplishment and the mastery of knowledge" (Miller and Mills, 1978, p. 10), but have not been subject to public scrutiny—at least, as Terry Hartle's essay points out, until now.

One reason the public and its representatives are now looking carefully at the credits and credentials awarded by institutions of higher education is that the Academy has made implicit claims for what they represent, similar to those made in the broader economy about licenses and certifications. In other words, we advance, as public and common, symbols that are privately (even idiosyncratically) defined. To the Academy, it is an article of faith that these symbols are measures of learning, yet the only public definition of both of them is couched in terms of time—and even then, in terms of time allocated for learning, not time actually used for learning. Furthermore, as Warren (1974) notes, the award of credit is an "all-or-none" situation that renders progress toward the degree a matter of mechanistic perseverance, and that the learning presumed to take place in most courses has rarely been validated by comprehensive assessment.

It is partly for this reason that our principal current interests in assessment in higher education lie in the cognitive dimensions of student growth, specifically in psychological (as opposed to behavioral) outcomes that take place during the college years and not afterwards. Ewell (1985) points out that these distinctions (psychological/behavioral; within-college/after-college; cognitive/affective) "combine and interact in many ways," and that the different combinations "define relatively distinct sets of research activities" that illustrate different aspects of "outcomes assessment." Some of these activities (cognitive, psychological, within-college) "require careful instrumentation" in pre/post-testing, others (affective, behavioral) call for surveys of students and alumni, and still others (behavioral, within-college) require the careful analysis of unobtrusive data such as "course-taking patterns, changes in student major and status, and retention (p. 3). But to put it simply, we are now concerned, more than ever, about publicly accessible knowledge of what students learn in college.

What Did the Papers in This Collection Do?

Recall that these papers were commissioned for a conference that a college catalogue might list as "Assessment 101." It was assumed that a majority of the 700 attendees at this conference were exploring the basics, and coming to terms with some fundamental questions, e.g. What are the ranges of methods and instruments available for different uses of assessment? How can assessment be used as an instructional tool? What are the costs of different types of program and institutional evaluations using assessment data? How does institutional type affect the objectives, development, and implementation of assessment programs? What is an "assessment center" and how does it work?

Appropriately enough, the papers do not provide the level of technical assistance that would be presented in "Assessment 301." Likewise, they do not approach assessment as principally a problem of institutional politics and faculty motivation, hence do not offer guidance on the implementation of innovations in academic organizations or on faculty development. Those are subjects for other conferences and other volumes, but not this one.

As a collection, the five papers reflect our current dichotomous view of traditional forms of testing and evaluation and emerging forms of performance assessment. The former dominate the presentations of Hartle, Harris, and Ewell/Jones, the latter of Loacker *et al.* and Elman/Lynton. The two views are conscious of each other. Each refers to the other. Each sometimes acknowledges that the other is appropriate, even successful, in certain circumstances. And yet, ultimately, each is skeptical of the other.

There *are* genuine differences between "production measures" and "recognition measures" (Cooper, 1984). The former require an individual to engage in an activity that directly embodies the desired skill or competence. The latter require an individual to judge simulated products of that activity. In the first, we observe behavior; in the second, a representation of behavior. These are two corner's of Webb's (1966) triangle for accurate measurement of any human activity: observation, trace, and archive.

The skepticism and dichotomous views are thus ultimately false. Assessment can (and does) use both types of measures. In both, instruments and methods are selected according to context, purpose, and practicality. Both are also subject to the canons of validity and reliability, and both rest on the principle of expert judgment.

Let's talk about a few of these concepts, not to provide a technical primer or even a preview of "Assessment 301," rather to insure that the reader reflects on these papers with a sense of some of the important structural features of assessment that they assume.

Validity. Whether we are talking about testing or performance, the current discussion of assessment in higher education cannot avoid the concept of validity. Validity is not a psychometrician's hocus-pocus: it is an absolute necessity in the structured judgment of human performance. The user of a method or instrument of assessment simply must be able to persuade others to accept the results according to the purposes of the assessment, and to accept what the results represent. There are, of course, different kinds of validity: predictive, content, and construct; but no assessment is immune to judgments of one or more of them. The closer an assessment methodology comes to the individual student's performance as the ultimate unit of analysis, the more predictive validity comes into play. Notice, when you read the Loacker *et al.* paper, that the developmental aspects of the Alverno approach to assessment essentially invite more judgments of predictive validity than some of the traditional testing approaches outlined by John Harris. At the same time, though, one of the virtues of the "Alverno doctrine" of multiple observation in assessment lies in its recognition that the predictive validity (or, as Mentkowski and Loacker (1985) call it, "performance validity") of a single measure decreases over time, particularly when the external criteria of performance change.

Nonetheless, predictive validity is very important in assessments such as basic skills placement tests administered to freshmen on entrance to college. A survey of over 600 community colleges by Woods (1985), for example, indicated that over half used systematic predictive validity research on placement tests (p. 22).

But some of the approaches advocated by Harris, particularly those involving essay or short-answer (as opposed to multiple-choice) examinations, often present problems in validity because the act of writing may interfere with other skills and knowledge being assessed. The problem is not insuperable, as Heywood (1977) has suggested, if we pay more attention "to the [content] validity of the question and the scheme [performance criteria] against which it is marked" (p. 36). In such assessments, attention also needs to be paid to the time frame. If speed of response is not a performance criterion for an essay examination (let alone for any other type of assessment), we dilute the validity of the method or instrument by placing arbitrary time constraints on students.

Expert Judgment and Reliability. Expert judgment is expressed in a number of ways in assessment, and there is no assessment in which the concept does not apply. Someone determines the content and standards of performance of every assessment. Whether those specifications are mushy or technically explicit, we oddly use a fallacy in argumentation—the appeal to authority—in tacitly accepting them. Of course, the setting and degree of acceptance varies. We demand reliability studies of standardized tests, the specifications for which are developed by experts in both content and psychometrics. Each of the Graduate Record Subject Area tests, for example, has a board of examiners appointed with the advice of the professional associations/learned societies in that field, and faculty in that field generally accept the expertise of these peer representatives in matters of setting the content specifications for the examinations. The development of each test is also conducted by psychometricians who can determine the difficulty-levels of questions, equate the scales of different versions of the same test, etc., and we generally accept their expert judgment in matters of performance scales and benchmarks. Once the reliability of those tests has been demonstrated, we seem relatively comfortable with the results.

In the realm of performance assessment, however, whether through essay examinations, simulations, etc., expert judgment involves a reliability problem (and a validity problem, as well, when one seeks to identify *external* expert judges for a performance assessment in a specific field). This has long been a criticism of classroom examinations designed and judged directly by individual faculty, and the suggestion is implicit in some of the papers in this volume (Harris, Loacker *et al.*, and Elman/Lynton) that the reliability of performance assessments would be enhanced by team development and multiple judgments of more than one expert. To the extent to which performance criteria are explicitly stated, to the extent to which there is consensus on the criteria, and to the extent to which the distinction among levels of performance is clear, then "inter-scorer" correlations can determine the reliability of an assessment. Where the correlations are comparatively high, say .65 or better, the task (and its performance criteria) can be retained. Where the correlations are lower, then one ought to reexamine both the task and the performance criteria.

Cooper (1984) also points out that reliability is an inherent problem in performance assessment "because different topics often require different skills or make different conceptual demands" on students (p. 4). While this is one more argument for multiple measures of student academic achievement in the same content or skill areas, the more general point is that reliability, like validity, applies not only to standardized testing but to other forms of assessment as well.

Criteria of Content and Performance

The new assessment movement in higher education prefers criterion-referenced measures to norm-referenced ("standardized") tests. While the dichotomy between the two is partially false, there is no question that criterion-referenced measures can serve more functions at the same time, provided that we can reach consensus on performance standards. It may be helpful to consider both halves of that sentence.

A criterion-referenced measure is designed to determine the degree of mastery of a body of knowledge or a skill by an individual student irrespective of the performance of other students. For that reason, the body of knowledge or skill or cognitive capacity (the "content domain") is defined in detail, and the definition is public, i.e. students know it, faculty know it, indeed, anyone who wants to know it can know it. A detailed analysis of the information concerning student mastery of the domains of content explicitly defined allows diagnostic uses of these measures for purposes of improving learning, instruction, and curriculum.

Norm-referenced measures also involve definitions of content domain (it's absolutely silly to claim that they don't), but those definitions tend to be more general and less public. One can infer these characteristics, for example, from content representativeness studies of various Graduate Record Examination Subject Area Tests (see Oltman, 1982; DeVore and McPeek, 1985). But norm-referenced measures are more concerned with comparing one student's performance to that of other students, and hence provide a different kind of information. This information is less useful for diagnostic purposes, but more useful for selection (Klitgaard, 1985).

If it is silly to claim that norm-referenced measures do not define a "content domain," it is equally silly to claim that criterion-referenced measures cannot be "standardized" and even normed. A long time ago, Ebel (1962) convincingly argued that to the extent to which we reach consensus on a domain of content, and generate equivalent tasks for students to demonstrate their mastery of that domain, we "standardize" a criterion-referenced measure. Who "reaches consensus?" If faculty do, then we can canonize continuous classroom assessment by aggregating judgments and raising them to the level of standards.

One is occasionally impressed with how well college professors can state the discrete competences, capacities, skills, and knowledge they expect students to develop. Then can, in fact, describe the *content* domain. But even in those cases there is a studious avoidance of performance criteria. They can tell us "what," but can't tell us how to recognize "how well." Such phrases as "evidences understanding," "demonstrates awareness," "communicates effectively," do not help anyone assess performance. There seems to be a limited and stock set of verbs that are mechanically generated in the process of writing criterion-referenced assessment tasks. Common sense suggests that the more limited and basic the vocabulary of performance, though, the less reliable the assessment.

It would not surprise me if a majority of college faculty found these statements awkward and childish. Indeed, a survey of departmental admissions committees and deans in graduate schools indicated that information on student performance presented in such forms—even in institutions receptive to nonstandard data—basically alienated them (Knapp and Hamilton, 1978).

Even in competency-based programs, criterion statements provide guidelines for *what* students are expected to do, not how well they are expected to perform. For example, an assignment "to evaluate the rhetorical effectiveness" of a communication "and its contribution to the effectiveness" of a communication "and its contribution to the effectiveness of the argument" in that communication, includes, as a performance standard, "evaluation of rhetorical effectiveness (25%)" (Hoyt, 1978, p. 144). Unfortunately, that type of tautological statement is more the rule than the exception. The exceptions, though, are worth noting, e.g. in an institutional program—Clayton Junior College's communication assessments—and in a testing program—the Academic Competences in General Education experiment conducted by Jonathan Warren in the late 1970s.

Developing statements of performance criteria that can be reliably applied by different faculty in different settings requires more work than most are willing to invest in the task. It is a matter of expanding our language space, of including a richness of verbs that describe what students do and do not do, and, more importantly, of using adverbs and using them well. Without the detailed standards that adverbs yield, the quality of information generated by an assessment suffers, and faculty are justly skeptical. So are public policymakers. In the absence of accuracy in criterion-referenced standards of performance, it becomes rather easy to turn to the certainty of a stanine.

There is nothing sophomoric about using the wealth of our language in establishing detailed, public criteria for both content domains and performance standards, teaching to them, and measuring student performance against them. As Secretary Bennett writes in the Foreword to this volume, "when a college or university does that . . . it simply does what it set out to do, and then checks to see how well it has succeeded." In this sense, he reminds us, there is nothing wrong or "shameful" about "teaching to the test."

Organizational and Policy Issues

There are a number of critical issues that current discussions concerning assessment gloss over, as if their mere mention causes discomfort. They ought to be noted here, so that discussions based on the work represented in this volume might take them into account.

The first concerns faculty resistance to third-party assessment. To some, assessment is a symbolic activity that says, in effect, we do not trust our faculty. If students see that faculty are not trusted, they will have one more reason for not pursuing academic careers. If faculty perceive that they are not trusted, it is said, we will have worse morale problems than we have already. The objection here is that the assessment of student learning—no matter what form it takes—will inevitably be used as an assessment of faculty competence. But this is like blaming the store that sold you the camera for the fuzzy pictures you took. It is the student who performs, not the faculty member.

One should note that faculty do not object to third-party assessment in matters of admissions and placement. Indeed, Woods' (1985) survey of community colleges reveals that the "primary source of pressure" to use tests in the admissions process is the faculty (p. 11).

The trust issue has another—and legitimate—dimension: the relationship of assessment to the promises we make to students. We promise to help students develop the intellectual capacities necessary to succeed in their careers and to live rich and rewarding lives. We promise to help them develop their writing and communication skills and capacities for reflective judgment. Yet if assessment promises *only* a mass of standardized, multiple-choice tests that rely principally on recognition, recall, and speed of response—none of which are higher order intellectual capacities, and all of which follow mechanical "fill in between the lines with a No. 2 pencil only please" formats—we will undercut all the efforts we otherwise make to improve writing, listening, and speaking, let alone to stimulate and enlarge the reflective capacities of the mind.

A second contentious and often ignored issue in these discussions is that of the investment of time demanded by assessment. As the paper by Ewell and Jones well demonstrates, assessment carried out for purposes of placement or program evaluation is not all that costly on a per-student basis. But some faculty can argue that, direct costs aside, they already devote an enormous amount of time to assessment, and that some of the new methodologies (e.g. those described by Loacker *et al.*) radically diminish the time allocated for instruction by replacing creative "enabling" activities with mechanical "certifying" activities. The argument that performance assessment is itself an instructional activity does not impress those who are already exhausted with careful reading and commentary on masses of papers and examinations, and who might say that a surfeit of assessment teaches the student a great deal about assessment but very little about anatomy and physiology, economic statistics, 19th century American fiction, or anything else students come to college to learn.

Third is the issue of the effects of assessment on minorities, particularly blacks and Hispanics. The common case is usually applied in discussions of tests designed and used for purposes of selection, e.g. the SATs or LSATs, where predictive validity is at issue, and in which blacks and Hispanics score significantly lower than whites and orientals; where the test scores tend to overpredict the subsequent academic performance of blacks in particular (Klitgaard, 1985); and where the preference of the critics is to change the tests rather than improve the education of these disadvantaged students.

Leaving that complex issue aside, however, the effects of assessment on minorities are, in fact, insidious—but for very different reasons than those presented in the common case. Simply by virtue of the politics of accountability that have created the competency-based basic skills programs in the urban school environments through which most of them pass, disadvantaged students are subject to a great deal of testing at the elementary and secondary level. The process of assessment, however, treats these students merely as vehicles for producing indicators of school performance, and teaches them so narrowly to the tests that they do not fully develop the type of learned abilities that are measured by the SATs or ACTs. Unfortunately, the basic skills centers at many colleges perpetuate this behavioristic instructional paradigm through programmed materials, and minorities tend to be disproportionally represented among the victims.

Fourth, assessment in higher education will not command either legitimacy or respect as long as it primarily seeks to certify comparatively low levels of cognitive skills. Given the realities of the political uses of language, "assessment" in *higher* education will carry negative symbolic baggage if it is perceived as insuring only that college graduates can utter grammatical sentences and perform basic arithmetic functions. The nature of our assessments express what society wants from higher education, and if that's what the assessments say, then eventually some state legislatures—let alone students, faculty, and administrators—will reject the methodology altogether.

In light of this issue, it is no wonder that elite institutions and flagship campuses of state universities are not leaders in the current assessment movement (the notable exceptions are principally liberal arts colleges such as Swarthmore and Hampshire, which have practiced rather creative approaches to assessment for decades). It has been observed, in fact, that the less selective the institution, the more likely faculty and administrators will seek to use assessment for purposes of instructional improvement and/or institutional development, marketing, and public relations. Whatever the benefits to those institutions (and they may be considerable), as long as the broader interest in assessment is confined to them, it will have little public credibility.

This exclusionary tendency is unfortunate for a number of reasons. First, complex institutions such as research universities have, within them, programs and professional schools that use an incredible variety of assessment methods, and as the paper in this collection by Elman and Lynton indicates, can be exemplary laboratories for the development and validation of these methods. Second, the fragmentation of all these methods and assessment activities within the complexity of a research university prevents institutional learning; but the literature on organizational structure and processes in universities suggests some very practical strategies for overcoming that fragmentation. If the most elite, complex, and influential institutions of higher education can demonstrate how much they can learn and improve by a coordinated assessment program, that learning will be more easily transferred to other institutions and to decisionmakers in state legislatures, state boards of higher education, and central system offices. Third, the faculty of these institutions are most likely to be members of committees that set the specifications for *de facto* national examinations such as the GREs and state licensure examinations in professional fields, so there is a natural base of experience in research universities with critical technical aspects of assessment such as defining subject domains, setting performance criteria, and determining the most reliable methods of administration. The more other faculty can learn from these, the greater the benefits to all institutions of higher education.

In Conclusion

What can we conclude from the papers in this volume and the issues raised in both Secretary Bennett's Foreword and this conclusion?

First, it is time for some serious study of assessment in American higher education by college faculty and administrators themselves. The intention of such study would not be to learn about assessment as an end in itself, rather, it would be to learn how to use assessment to improve curriculum and instruction, and as an occasion for reflecting on both what it means to be educated at the college level in individual disciplines and what it means to develop the various cognitive capacities of young adults.

Secondly, it is also time for *critical* analysis. We've witnessed too much blind enthusiasm in some quarters, and deaf rejections in others. For there to be critical analysis, at least a modicum of technical knowledge is necessary. There are those too eager to emulate the "value-added" model at Northeast Missouri State, the developmental model at Alverno, or the comprehensive performance model as practiced in assessment centers run by major employers. There are significant problems with each of these models irrespective of issues concerning organizational context and transferrability. Klitgaard's (1985) discussion of the difficulties of translating the "theoretically attractive" notion of value-added into measurements useful in an imperfect society and an ambiguous future is well worth pondering as an example. If we are serious at all about improving the education of college students, and using assessment as one of our tools, then we cannot gloss over these problems.

Lastly, the hour for polemics is over. Addressing the American Association for Higher Education in 1980, Francis Keppel contended that "despite the rhetoric and generalizations that we have all used, we do not have the kind of detailed and comparable information on student performance" that enables students, faculty, institutions, accrediting associations, and state governments "to make the choices" that each party has to make to participate effectively in a system founded on human judgment. Six years and as many national reports later, we are just starting to develop that information. The parties owe it to each other to drop the polemics and get to work.

References

Not all these bibliographical entries are cited in the above text, nor, with few exceptions, are they cited elsewhere in this volume. They are listed in the spirit of assistance, as I have found them helpful in studying and thinking about assessment over the past few years.

Anastasi, A. *Psychological Testing*. 4th edition. New York: Macmillan, 1976.

Astin, A. W., "Undergraduate Achievement and Institutional Excellence." *Science*, vol. 161, no. 3842 (1968), pp. 661-668.

Astin, A. W., "The Methodology of Research on College Impact: Part One." *Sociology of Education*, vol. 43 (1970), pp. 223-254.

Berk, R. A. *Criterion-Referenced Measurement: The State of the Art*. Baltimore: Johns Hopkins University, 1980.

Blackburn, R. *et al. Changing Practices in Undergraduate Education*. Berkeley: Carnegie Council on Policy Studies in Higher Education, 1977.

Bowen, H. *Investment in Learning*. San Francisco: Jossey-Bass, 1977.

Brown, D., "A Scheme for Measuring the Output of Higher Education," in Lawrence, B., Weathersby, G. and Patterson, V.W. (eds.), *Outputs in Higher Education: Their Identification, Measurement and Evaluation*. Boulder, CO: Western Interstate Commission for Higher Education, 1970, pp. 27-40. ED#043-296.

Buros, O., "Fifty Years in Testing: Some Reminiscences, Criticisms and Suggestions." *Educational Researcher*, vol. 6, no. 7 (1977), pp. 9-15.

Coffman, W. E., "On the Validity of Essay Tests of Achievement." *Journal of Educational Measurement*, vol. 3 (1966), no. 2, pp. 151-156.

Conrad, L., Trisman, D. and Miller, R. (eds.). *Graduate Record Examination Technical Manual*. Princeton: ETS, 1977.

Cooper, P. L. *The Assessment of Writing Ability: A Review of the Research*. GREB No. 82-15R. Princeton: ETS, 1984.

Cronbach, L. J. and Furby, L., "How Should We Measure 'Change'—or Should We?" *Psychological Bulletin*, vol. 74 (1970), pp. 68-80.

Cronbach, L. J., "Five Decades of Public Controvery Over Mental Testing," *American Psychologist*, January 1975, pp. 1-14.

Cummins, J., "The Cross-Lingual Dimensions of Language Proficiency," *TESOL Quarterly*, vol. 4 (1980), pp. 175-187.

DeVore, R. and McPeek, M. *Report of a Study of the Content of Three GRE Advanced Tests*. GREB No. 78-4R. Princeton: ETS, 1985.

Ebel, R. L., "Content-Standard Test Scores." *Educational and Psychological Measurement*, vol. 22 (1962), pp. 15-25.

Edmonds, R. R., "Testing and Educational Equity: The Status Quo and Prospects for the Future," in Schrader, W. B. (ed.), *Measurement, Guidance and Program Improvement*. San Francisco: Jossey-Bass, 1982, pp. 5-12.

Ewell, P. T. (ed.) *Assessing Educational Outcomes*. New Directions in Institutional Research No. 47. San Francisco: Jossey-Bass, 1985.

Ewens, T., "Analyzing the Impact of Competence-Based Approaches on Liberal Education," in G. Grant and Associates, *On Competence*. San Francisco: Jossey-Bass, 1979, pp. 160-198.

Ferguson, R. L., "Assessing Learning for Credit and Credentials," in Miller, J. W. and Mills, O. (eds.), *Credentialing Educational Accomplishment*. Washington: American Council on Education, 1978, pp. 95-129.

Gardner, E., "Some Aspects of the Use and Misuse of Standardized Aptitude and Achievement Tests," in Wigdor, A. K. and Garner, W. R. (eds.), *Ability Testing: Uses, Consequences and Controversies*. Washington: National Academy of Sciences, 1982, vol. II, pp. 315-332.

Harris, C. W., Alkin, M. C. and Popham, W. J. (eds.). *Problems in Criterion-Referenced Measurement*. Los Angeles: UCLA, 1974.

Hartnett, R. T. and Centra, J. A. "The Effects of Academic Departments on Student Learning." *Journal of Higher Education*, vol. 48 (1977), pp. 491-507.

Heywood, J. *Assessment in Higher Education*. London: John Wiley and Sons, 1977.

Hoyt, D., "Competence as a Basis for Credit and Credentialing," in Miller, J. W. and Mills, O. (eds.), *Credentialing Educational Accomplishment*. Washington: American Council on Education, 1978, pp. 130-158.

Keeley, S., Browne, M. and Kreutzer, J., "A Comparison of Freshmen and Seniors on General and Specific Essay Tests of Critical Thinking." *Research in Higher Education*, vol. 17 (1982), pp. 139-154.

Keppel, F. "Basic Public Policy on Student Achievement: A New Agenda Item." Keynote Address, National Conference on Higher Education, American Association for Higher Education, Washington, 1980.

Klitgaard, R. *Choosing Elites*. New York: Basic Books, 1985.

Knapp, J. and Hamilton, I. B. *The Effect of Nonstandard Undergraduate Assessment and Reporting Practices on the Graduate School Admissions Process*. GREB No. 76-14R. Princeton: ETS, 1978.

Kruskal, W., "Formulas, Numbers, Words: Statistics in Prose." *The American Scholar*, vol. 47, no. 2 (1978), pp. 223-229.

Mehrens, W. A. and Ebel, R. L., "Some Comments on Criterion-Referenced and Norm-Referenced Achievement Tests." *Measurement in Education*, vol. 10, no. 1 (1979). pp. 1-7.

Mentkowski, M. and Loacker, G. "Assessing and Validating the Outcomes of College," in Ewell (1985), pp. 48-64.

Miller, J. W. and Mills, O. (eds.) *Credentialing Educational Accomplishment: Report and Recommendations of the Task Force on Educational Credit and Credentials*. Washington: American Council on Education, 1978.

Oltman, P. K. *Content Representativeness of the GRE Advanced Tests in Chemistry, Computer Science and Education*. GREB 81-12p. Princeton: ETS, 1982.

Pace, C. R. *Achievement and the Quality of Student Effort*. Paper prepared for the National Commission on Excellence in Education, ED# 227-101, 1982.

Pascarella, E., "College Environmental Influences on Learning and Cognitive Development: A Critical Review and Synthesis," in Smart, J. (ed.), *Higher Education: Handbook of Theory and Research*. New York: Agathon Press and AERA, Division J, 1985, pp. 1-61.

Piaget, J., "Intellectual Development from Adolescence to Adulthood." *Human Development*, vol. 15 (1972), pp. 1-12.

Swinton, S. S. and Powers, D. E. *A Study of the Effects of Special Preparation on GRE Analytical Scores and Item Types*. GREB no. 78-2R. Princeton: ETS, 1982.

Terenzini, P., Pascarella, E., and Lorang, W., "An Assessment of the Academic and Social Influences on Freshman Year Educational Outcomes." *Review of Higher Education*, vol. 5 (1982), pp. 86-109.

Vaizey, J., "The Outputs of Higher Education: Their Proxies, Measurement, and Evaluation," in Lawrence, B., Weathersby, G. and Patterson, V. W. (eds.), *Outputs of Higher Education: Their Identification, Measurement, and Evaluation*. Boulder, CO: Western Interstate Commission for Higher Education, 1970, pp. 19-26. ED#043-296.

Warren, J. R., "Awarding Credit," in Cross, K. P., Valley, J. R. and Associates, *Planning Non-Traditional Programs*. San Francisco: Jossey-Bass, 1974, pp. 116-147.

Warren, J. R. *The Measurement of Academic Competence*. Final Report, Grant #G007603526. Washington, DC: Fund for the Improvement of Postsecondary Education, 1978.

Webb, E. *et al. Unobtrusive Measures*. Chicago: Rand-McNally, 1966.

Wigdor, A. K. and Garner, W. R. (eds.) *Ability Testing: Uses, Consequences, and Controversies*. 2 vols. Washington: National Academy of Sciences, 1982.

Wilson, K. M. *The Relationship of GRE General Test Item-Type Part Scores to Undergraduate Grades*. GREB no. 81-22p. Princeton: ETS, 1985.

Witkin, N. A. *et al. A Longitudinal Study of the Role of Cognitive Styles in Academic Evolution During the College Years*. GREB no. 76-10R. Princeton: ETS, 1977.

Woods, J. *Status of Testing Practices at Two-Year Postsecondary Institutions*. Iowa City: American College Testing Service and the American Association of Community and Junior Colleges, 1985.

Note: It is unfortunate that too many of these entries can be classified as "fugitive literature." The most indispensible and accessible of the items cited above would be Wigdor and Garner, Pascarella, and Heywood.